# Genotyping-by-sequencing illuminates high levels of divergence among sympatric forms of coregonines in the Laurentian Great Lakes

Amanda S. Ackiss[1], Wesley A. Larson[2], Wendylee Stott[3]

[1] Wisconsin Cooperative Fishery Research Unit, College of Natural Resources, University of Wisconsin-Stevens Point, 800 Reserve St., Stevens Point, WI 54481, aackiss@uwsp.edu

[2] U.S. Geological Survey, Wisconsin Cooperative Fishery Research Unit, College of Natural Resources, University of Wisconsin-Stevens Point, 800 Reserve St., Stevens Point, WI 54481, wlarson@uwsp.edu

[3] U.S. Geological Survey, Great Lakes Science Center, 1451 Green Rd., Ann Arbor, MI 48105, wstott@contractor.usgs.gov

***Corresponding author**

Amanda S. Ackiss

Mailing address:
College of Natural Resources
University of Wisconsin Stevens Point
800 Reserve St.
Stevens Points, WI 54481-3897 USA

Email: aackiss@uwsp.edu

**Acknowledgements**

We thank Brad Ray with the Wisconsin Department of Natural Resources and Dan Yule and Mark Vinson with the United States Geological Survey for providing the Lake Superior samples used in this project. We are grateful to Kristen Gruenthal for laboratory support during RAD library preparation, to Garrett McKinney for discussions on analysis, and to Julie Turgeon, John Robinson, and Jared Homola for comments on the manuscript. This research was supported by the Turing High Performance Computing cluster at Old Dominion University and was funded by grants from the Great Lakes Restoration Initiative, the Great Lakes Fishery Trust, and the Great Lakes Fish and Wildlife Restoration Act. Any use of trade, product, or company name is for descriptive purposes only and does not imply endorsement by the U.S. Government.

## Abstract

Effective resource management depends on our ability to partition diversity into biologically meaningful units. Recent evolutionary divergence, however, can often lead to ambiguity in morphological and genetic differentiation, complicating the delineation of valid conservation units. Such is the case with the "coregonine problem," where recent post-glacial radiations of coregonines into lacustrine habitats resulted in the evolution of numerous species flocks, often with ambiguous taxonomy. The application of genomics methods is beginning to shed light on this problem and the evolutionary mechanisms underlying divergence in these ecologically and economically important fishes. Here, we used restriction site-associated DNA (RAD) sequencing to examine genetic diversity and differentiation among sympatric species in the *Coregonus artedi* complex in the Apostle Islands of Lake Superior, the largest lake in the Laurentian Great Lakes. Using 29,068 SNPs, we were not only able to clearly distinguish the three most common forms for the first time, but putative hybrids and potentially mis-identified specimens as well. Assignment rates to form with our RAD data were 93-100% with the only mis-assignments arising from putative F1 hybrids, an improvement from 62-77% using microsatellites. Estimates of pairwise differentiation ($F_{ST}$: 0.045-0.056) were large given the detection of hybrids, suggesting that hybridization among forms may not be successful beyond the F1 state. We also used a newly built *C. artedi* linkage map to look for islands of adaptive genetic divergence among forms and found widespread differentiation across the genome, a pattern indicative of long-term drift, suggesting that these forms have been reproductively isolated for a substantial amount of time. The results of this study provide valuable information that can be applied to develop well-informed management strategies and stress the importance of re-evaluating conservation units with genomic tools to ensure they accurately reflect species diversity.

**Keywords:** adaptive divergence, management units, genomic islands of divergence, species complex, RAD sequencing, coregonines, population genomics, hybridization

## Introduction

Defining conservation units is one of the most fundamental yet challenging aspects of resource management (Coates, Byrne, & Moritz, 2018). Partitioning species into units with substantial reproductive isolation provides managers with the ability to monitor and regulate independently evolving groups that may respond differently to harvest, disease, habitat alteration, or climate change (Allendorf & Luikart, 2007; Ryder, 1986). Over the past several decades, advancements in genetic analysis have provided scientists with powerful tools to estimate the amount of gene flow between species or populations to inform the creation of conservation units (Olsen et al., 2014; Palsbøll, Bérubé, & Allendorf, 2007; Palsbøll, Peery, & Bérubé, 2010; Schwartz, Luikart, & Waples, 2006). With the arrival of the genomics era, the power and accuracy to discern levels of reproductive isolation, inbreeding, and effective population size has vastly improved (Allendorf, Hohenlohe, & Luikart, 2010), and tools such as genome scans have revolutionized our ability to identify and understand adaptative genetic variation (Funk, McKay, Hohenlohe, & Allendorf, 2012; Waples & Lindley, 2018). Despite these advancements, delineating discrete conservation units can still be problematic. For example, taxonomic uncertainty can lead to confusion regarding species boundaries (Bickford et al., 2007; Hey, Waples, Arnold, Butlin, & Harrison, 2003), and observed phenotypic, spatial, temporal, or behavioral differences can be opposed by apparent genetic panmixia (Als et al., 2011; Hoey & Pinsky, 2018; Palm, Dannewitz, Prestegaard, & Wickström, 2009).

Perhaps no other group embodies the challenges of defining conservation units better than the coregonines. A subfamily of the Salmonidae, coregonines are comprised of three genera of freshwater and anadromous fishes distributed throughout cold water habitat in North America, Europe, and Asia. The most speciose genus, *Coregonus*, includes the ciscoes and whitefishes, which exhibit an extreme array of phenotype variability that is attributed to recent adaptive radiation into lacustrine habitat following glacial retreat during the Pleistocene epoch (Schluter, 1996). Often, distinct phenotypes can be found both in sympatry and allopatry which leads to difficulty in distinguishing a single, monophyletic origin of forms

89 from parallel ecological speciation in individual lakes. Several coregonines exhibit sympatric dwarf and
90 normal forms, including European whitefish *C. lavaretus*, North American whitefish *C. clupeaformis*,
91 cisco *C. artedi*, and least cisco *C. sardinella* (Huitfeldt-Kaas, 1918; Mann & McCart, 1981; Shields,
92 Guise, & Underhill, 1990; Vuorinen, Bodaly, Reist, Bernatchez, & Dodson, 1993). Empirical evidence of
93 hybridization and introgression (Garside & Christie, 1962; Kahilainen et al., 2011; Lu, Basley, &
94 Bernatchez, 2001) raises the question of how to manage forms when reproductive isolation is incomplete
95 and has led some to suggest that the broad phenotypic variation observed in coregonines is a result of
96 reticulate evolution (Svärdson, 1998; Turgeon & Bernatchez, 2003). This breadth of taxonomic ambiguity
97 in the coregonines was first termed "the coregonid problem" by Svärdson (1949) and persists today as the
98 more accurate "coregonine problem" (Eshenroder et al., 2016; Mee, Bernatchez, Reist, Rogers, & Taylor,
99 2015).

100 In recent decades phenotypic data have been combined with genetic analysis in an attempt to untangle the
101 complicated relationships among coregonines. Taxonomic units in coregonine systematics have
102 traditionally relied on morphological characteristics such as head and body shape, morphometrics, and
103 meristics such as gill raker counts (Himberg, 1970; Koelz, 1929; Svärdson, 1979). Phylogeographic
104 analyses with allozymes, restriction fragment length polymorphisms (RLFPs), mitochondrial DNA
105 (mtDNA), and microsatellites have helped resolve evolutionary relationships among species and forms
106 (Bernatchez & Dodson, 1994; Østbye, Bernatchez, Naesje, Himberg, & Hindar, 2005), and this growing
107 body of research indicates that relying solely on phenotypic traits can be problematic for determining
108 phylogenetic relationships when environmental plasticity occurs (e.g. Muir et al., 2013; Todd, 1998;
109 Todd, Smith, & Cable, 1981). Amplified fragment length polymorphisms (AFLPs), mtDNA, and
110 microsatellite analyses all show support for broad colonization of monophyletic lineages followed by the
111 parallel ecological speciation of sympatric forms of European whitefish *C. lavaretus* (Hudson, 2011;
112 Hudson, Lundsgaard-Hansen, Lucek, Vonlanthen, & Seehausen, 2017; Præbel et al., 2013). Recently,
113 genomics methods have been applied to the coregonine problem in lake whitefish *C. clupeaformis*, and

114 restriction site-associated DNA (RAD) sequencing, quantitative trait loci (QTL) analysis, and genome
115 scans have provided valuable insight into the evolutionary mechanisms of speciation-with-gene-flow in
116 dwarf and normal forms (Gagnaire, Normandeau, Pavey, & Bernatchez, 2013; Gagnaire, Pavey,
117 Normandeau, & Bernatchez, 2013; Laporte et al., 2015; Rogers & Bernatchez, 2007).

118 The most extensive regional adaptive radiation within the *Coregonus* genus in North America occurred in
119 the Laurentian Great Lakes, but detection of genetic differentiation or reproductive isolation among Great
120 Lakes forms has been mostly unsuccessful. Rapid diversification of the shallow water cisco *C. artedi*
121 (Eshenroder et al., 2016; Turgeon & Bernatchez, 2003) into newly available deepwater habitat following
122 the Wisconsin Glacial Episode resulted in the evolution of at least eight distinct forms (Koelz, 1929; Scott
123 & Crossman, 1998). Morphological differences occur across a variety of traits including body and head
124 shape, lower jaw position, eye size, fin length, and gill raker counts (Koelz, 1929), though subtle
125 variations among forms and between lakes can often make visually distinguishing them difficult without
126 all possible forms present (Eshenroder et al., 2016; Turgeon et al., 2016). Different forms typically occur
127 in specific depth ranges, and stable isotope analysis supports niche differentiation indicating that many of
128 the Great Lakes forms occupy different trophic levels with observed changes in proportion of pelagic and
129 benthic food sources (Schmidt, Harvey, & Vander Zanden, 2011; Schmidt, Vander Zanden, & Kitchell,
130 2009; Sierszen et al., 2014). All forms likely undergo seasonal spawning migrations, forming nearshore
131 aggregations in November and December (Stockwell, Hrabik, Jensen, Yule, & Balge, 2010; Yule,
132 Stockwell, Evrard, Cholwek, & Cullis, 2006; Yule, Addison, Evrard, Cullis, & Cholwek, 2009).
133 However, very little is known about behavioral differences between forms during overlapping periods of
134 spawning, maintaining the possibility that hybridization during spawning events could be preventing
135 genetic divergence. Analyses using RFLPs, mtDNA, and microsatellites have resulted in estimates of low
136 or no genetic differentiation among Great Lakes forms (Bernatchez, Colombani, & Dodson, 1991; Reed,
137 Dorschner, Todd, & Phillips, 1998; Turgeon & Bernatchez, 2003; Turgeon et al., 2016) leaving the
138 question of reproductive isolation - particularly among deepwater forms – unanswered.

Over the past century, anthropogenic impacts have greatly reduced the original diversity of the *C. artedi* species complex in the Great Lakes, underscoring the need for establishing well-informed conservation units in ciscoes. The introduction of invasive forage fish, overfishing, and habitat loss led to large decreases in cisco abundance and lake-wide extirpation to complete extinction of historically documented deepwater forms (Commission & Christie, 1973; Smith, 1968, 1970; Wells & McClain, 1973). Of the eight accepted forms originally described, only *C. artedi* and three deepwater forms – *C. hoyi*, *C. kiyi*, and *C. zenithicus* - are extant in the Great Lakes (referred to henceforth by specific epithet; Bailey and Smith, 1981; Todd and Smith, 1992). A fifth deepwater form, *C. nigripinnis,* has been reduced to nearby Lake Nipigon (Ontario, Canada), though an extant *nigripinnis*-like form is still periodically caught in Lake Superior (Eshenroder et al., 2016). Lake Superior's peripheral location relative to both large human populations (and associated fishing pressure) and the canal construction that opened the Great Lakes to invasion from non-native species appears to have provided some protection from the impacts that extirpated cisco forms from the other four lakes (Koelz, 1926). Of the four lakes in which members of the *C. artedi* complex remain, Lake Superior is the only lake where all extant Great Lakes forms can still be regularly found (Eshenroder et al., 2016).

Recent evidence in the Great Lakes for declining abundance of invasive fish such as alewife and increasing abundance in cisco (Bronte et al. 2003, Mohr and Ebener 2005, Schaeffer et al. 2007) has led to growing interest in re-establishing lost populations. An understanding of the roles of heritable genetic differences and reproductive isolation in the establishment and persistence of remnant forms is vital for developing both informed conservation units and restoration strategies. The main goal of our study was to employ genomic methods to improve our understanding of genetic variation among these forms. Specifically, we (1) examined genetic differentiation and diversity among putative forms of cisco (2) compared the performance of SNPs and microsatellites in this system, and (3) leveraged a newly built cisco linkage map (Blumstein, 2019) to investigate adaptive divergence among forms. In order to remove the potentially confounding factor of distinguishing spatial genetic structure from form-based genetic

164 structure, we focused on a single region in Lake Superior where multiple forms of cisco are found in
165 sympatry, the Apostle Islands.

166 **Materials and Methods**

167 Tissue samples preserved in >95% ethanol were collected for the three most common cisco forms– *artedi*,
168 *hoyi*, and *kiyi* in the Apostle Islands (Fig. 1, Table 1). Samples of putative *artedi* were collected in
169 November 2017 using top and bottom gillnet surveys off Madeline Island conducted by the Wisconsin
170 Department of Natural Resources. Putative *artedi*, *hoyi* and *kiyi* were collected in July 2005 off Stockton
171 Island by the U.S. Geological Survey Great Lakes Science Center (GLSC). In addition, we included a
172 small number of available samples from two rare forms of cisco – *zenithicus* and the *nigripinnis*-like –
173 from Minnesota, Michigan, and Ontario waters in Lake Superior collected by the GLSC in summer 2007,
174 2012-2015. All samples were identified to putative form using a suite of standardized morphological
175 characteristics (Eshenroder et al., 2016). DNA was isolated from fin tissue samples with Qiagen
176 DNeasy® Blood & Tissue Kits.

177 *RAD library prep and sequencing*

178 Restriction site-associated DNA (RAD) libraries were prepared following the BestRAD protocol (Ali et
179 al., 2016). Extracted DNA was quantified using a Quant-it™ PicoGreen® dsDNA Assay (Invitrogen,
180 Waltham, MA) and normalized to a concentration of approximately 50 ng/µl for a 2 µl digestion reaction
181 with the restriction enzyme *SbfI* followed by ligation with barcoded adaptors. Individually barcoded
182 libraries were pooled into master libraries of 96 and fragmented to ~300-500bp with 12-14 30s cycles in a
183 Q500 sonicator (Qsonica, Newtown, CT). Fragmented DNA containing library adapters was bound to
184 Dynabeads™ M-280 Streptavidin magnetic beads (Invitrogen) and washed with buffer to remove non-
185 target fragments before an incubation step to release DNA from the beads. Following purification with
186 AMPure XP beads (Beckman Coulter, Brea, CA) master libraries were input into the NEBNext® Ultra™
187 DNA Library Prep Kit for Illumina® at the End Prep step for ligation of master library barcodes, a 250-

bp insert size-selection, and a 12-cycle PCR enrichment. Successful size-selection and enrichment were confirmed with visualization of products on a 2% agarose E-Gel (Invitrogen). Products underwent a final AMPure XP purification clean-up followed by quantification with a Qubit® 2.0 Fluorometer. All prepared libraries were sent to Novogene (Sacramento, CA) for sequencing on the Illumina NovaseqS4 platform.

*Read processing and SNP filtering*

Raw sequences generated from RAD sequencing were processed in the software pipeline Stacks v2.3d (Catchen, Amores, Hohenlohe, Cresko, & Postlethwait, 2011; Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013). Sequences were demultiplexed by barcode, filtered for presence of the enzyme cut-site and quality, and trimmed in the subprogram *process_radtags* (parameter flags: -e *SbfI* -c -q -r -t 140 --filter_illumina --bestrad). Filtered reads for each individual were aligned to create matching stacks with *ustacks* following guidelines suggested from empirical testing to avoid under- or over-merging loci from RAD datasets (Paris, Stevens, Catchen, & Johnston, 2017; parameter flags: --disable-gapped -m 3 -M 5 -H --max_locus_stacks 4 --model_type bounded --bound_high 0.05). A catalog of consensus loci built from 51 *artedi* sampled in Lake Huron used for the development of a cisco linkage map (Blumstein, 2019) was appended in *cstacks* with an additional 75 individuals in the *C. artedi* species complex from across the Great Lakes, including 19 fish used in the current study: five *hoyi* and four *kiyi* from the Apostle Islands and five each of *nigripinnis*-like and *zenithicus* from various locations across Lake Superior (parameter flags: -n 3 -p 6 –disable_gapped). Locus stacks for each individual were matched to the catalog using *sstacks* (parameter flag: --disable_gapped), data were oriented by locus in *tsv2bam*, and reads were aligned to loci and SNPs were called with *gstacks*. SNPs genotyped in greater than 30% of individuals (parameter flag: -r 0.3) were exported with the subprogram *populations* in genepop and variant call format (vcf) files.

Primary SNP filtering was performed with vcftools v0.1.15 (Danecek et al., 2011) and included (1) removing loci genotyped in fewer than 80% of individuals, (2) removing individuals missing more than

50% of loci, and (3) removing loci with a minor allele count less than 3. In addition, since all salmonids including members of the *C. artedi* species complex have experienced a recent genome duplication, putatively paralogous loci were identified with the program HDPlot (McKinney, Waples, Seeb, & Seeb, 2017), and any loci with heterozygosity greater than 0.55 or a read ratio deviation greater than 5 and less than –5 were removed. Finally, loci on the same RAD tag may be linked so only the SNP with the highest minor allele frequency on each tag was included in the final dataset. All file format conversions were performed using PGDSpider v2.1.1.5 (Lischer & Excoffier, 2011).

*Microsatellite amplification and genotyping*

We used the methods described in Stott et al. (in press) to genotype 12 microsatellites developed for coregonines (Bernatchez, 1996; Patton, Gallaway, Fechhelm, & Cronin, 1997; Rogers, Marchand, & Bernatchez, 2004) and salmonids (Angers, Bernatchez, Angers, & Desgroseillers, 1995; Estoup, Presa, Krieg, Vaiman, & Guyomard, 1993) in *artedi*, *hoyi*, and *kiyi* from the Apostle Islands: *Bwf*1, *Bwf2*, *C2*-157, *Cocl*23, *CoclLav*6, *CoclLav*27, *CoclLav*32, *CoclLav*72, *Sfo*8, *Sfo*23, *Str*-60. *Sfo8* consistently amplifies two genomic regions resulting in alleles sorting into upper (U: 215-281bp) and lower (L: 163-193bp) size ranges, therefore *Sfo8* is considered two loci. Fragment analysis was performed using a Genetic Analyzer 3.0 (Life Technologies), and genotypes were assigned at each locus using GeneMapper 3.7 (Life Technologies). We used Genepop v4 (Rousset, 2008) to conduct exact tests for deviations from Hardy–Weinberg and linkage equilibrium ($\alpha = 0.01$). Three loci were removed for being out of Hardy-Weinberg Equilibrium in all three forms (*Bwf*1, *Cocl*23, *CoclLav*6), and no loci showed significant linkage disequilibrium.

*Genetic differentiation and diversity*

To ensure that our putative form designations were appropriate, we used two approaches to assess genetic similarity among individuals. First, we conducted a principal component analysis (PCA) in the R package 'adegenet' (Jombart, 2008) for both the RAD and microsatellite datasets. Next, we estimated the number

of ancestral populations, *K*, contributing to contemporary genetic clustering for the RAD dataset using the program ADMIXTURE v1.3 (Alexander, Novembre, & Lange, 2009). We tested *K* from 1-5 with ADMIXTURE's cross-validation procedure and a k-fold of 10 (parameter flag: --cv=10) to examine support for each *K*. A *Q*-score of less than 70% was used to identify putative hybrids following similar thresholds applied in the literature (Kapfer, Sloss, Schuurman, Paloski, & Lorch, 2013; Marie, Bernatchez, & Garant, 2011; Weigel et al., 2018), and hybrid combinations were assigned to the two populations representing the largest *Q*-scores within the hybrid individual. ADMIXTURE analysis was only conducted on the RAD dataset because this program is not compatible with microsatellite data.

PCA and ADMIXTURE analysis with RAD genotypes revealed that three individuals originally identified as *hoyi* fell within the cluster of *kiyi* samples and, given the discrete clustering of *kiyi* and the strong possibility that these three *hoyi* samples were misidentified in the field, we removed these individuals from both RAD and microsatellite datasets for all further analyses to prevent bias in estimates of diversity, differentiation, and $N_e$.

We calculated a variety of summary statistics for the groups comprised of the five forms (*artedi* -ART, *hoyi* – HOY, *kiyi* – KIY, *nigripinnis* – NIG, and *zenithicus* – ZEN) including percentage of total observed alleles, observed and expected heterozygosity, and an inbreeding coefficient ($G_{IS}$). Summary statistics for each locus and populations were calculated using both microsatellite and RAD datasets in GenoDive v2.0b23 (Meirmans & Van Tienderen, 2004). Genetic differentiation among all forms was estimated across all loci in the RAD dataset with pairwise $F_{ST}$ (Weir & Cockerham, 1984) in Genepop and tested using exact tests (Goudet et al., 1996; Raymond and Rousset, 1995; alpha = 0.01) in Arlequin (Excoffier & Lischer, 2010). We also calculated locus-specific overall and pairwise-$F_{ST}$ values (Weir & Cockerham, 1984) in Genepop using a dataset that included the three forms with n>10 (ART, HOY, and KIY). To compare genetic differentiation between RAD and microsatellite datasets, we used GenoDive to estimate standardized pairwise genetic differentiation, $G'_{ST}$ (Hedrick, 2005), which employs an additional correction for bias from sampling a limited number of populations (Meirmans & Hedrick, 2011).

The rate at which individuals were able to be assigned back to their form with both RAD and microsatellite datasets was tested using population assignment in GenoDive for all forms with n>10. Assignment was performed by calculating the home likelihood ($L_h$) that an individual genotype is from a specific group given the allele frequencies (Paetkau, Calvert, Stirling, & Strobeck, 1995) using the leave-one-out method to avoid the bias from a target individual's contribution to the allele frequencies of a source population. Zero frequencies were replaced with 0.005 and a significance threshold of alpha=0.002 was applied separately to each group over 1000 replicated datasets (c.f. Perreault-Payette et al., 2017).

Effective population size ($N_e$) was estimated for all forms with n>10 using both RAD and microsatellite datasets with the bias-corrected linkage disequilibrium method (LDNE; Hill, 1981; Waples, 2006; Waples & Do, 2010) in the software package NeEstimator v2.1 (Do et al., 2014). We used a p-crit of 0.05 for the RAD dataset (Waples, Larson, & Waples, 2016) and 0.02 for the microsatellite dataset (Waples & Do, 2010). For the RAD dataset, only comparisons between sampled loci that were found on different linkage groups (LGs) of the cisco linkage map (Blumstein, 2019) were included to correct for physical linkage (Waples et al., 2016). $N_e$ calculations using the linkage disequilibrium method can be biased slightly downward when individuals from multiple cohorts are included in the sample due to a slight Wahlund effect (7% downward bias on average; Waples, Antao, & Luikart, 2014). However, this small bias should not greatly affect the interpretation of the $N_e$ results.

Finally, we used EASYPOPv2.0.1 (Balloux, 2001) to compare observed patterns of differentiation and hybridization with simulations encompassing a variety of demographic scenarios. Base parameters for all simulations were informed where possible by cisco life history traits and included random mating, same number of females and males per population (total n=2,000 per population based on our $N_e$ estimates for the three main cisco forms), equal proportions of female and male migration, biallelic loci with free recombination, a mutation rate equivalent to that measured in human and fish nuclear genomes ($\mu=1.0 \times 10^{-8}$; Bernardi & Lape, 2005; Conrad et al., 2011) and the KAM mutation model. First, since successful hybridization among forms will result in introgression and decreased differentiation, we modeled the

impacts of hybridization on genetic differentiation with three scenarios: two populations with a starting level of differentiation equal to 1) approximately the level we observed amongst cisco forms with our RAD dataset ($F_{ST}$ ≈0.05), 2) approximately twice the amount observed ($F_{ST}$ ≈0.10), and 3) approximately four times the amount observed ($F_{ST}$ ≈0.20). Preliminary levels of differentiation were achieved with low migration rates (m) set over 1,000 generations (m=0.001, m=0.0005, and m=0.0001, respectively) and followed by 5, 10, or 15 generations of one of three higher migration (i.e. hybridization) rates (m=0.01, 0.05, and 0.10) for every initial level of differentiation. Each unique parameter combination was run using a dataset of 1,000 loci and replicated 50 times. Estimates of $F_{ST}$ for each replicate were generated in the R package 'diveRsity' (Keenan, McGinnity, Cross, Crozier, & Prodöhl, 2013) with a genepop file containing a subset of 100 randomly selected individuals per population (50 females/50 males). Second, we simulated migration between three populations representing our three common cisco forms (ART, HOY, KIY) in order to reconstruct datasets with similar characteristics to our empirical data and compare signatures of hybridization among them. Preliminary differentiation among the three populations (P1-P3) was set at $F_{ST}$ ≈0.05 using the same method described above, and we allowed populations to hybridize for 2, 5, or 10 generations with m=0.05, which was equal to roughly the proportion of hybrids observed in our wild populations between ART-HOY and HOY-KIY. Since no putative ART-KIY hybrids were observed in our RAD dataset, we chose to implement a one-dimensional stepping stone model of migration (Kimura & Weiss, 1964) among simulated populations. Each unique parameter set for these simulations was run using 5,000 loci, a genepop file was output containing a subset of 100 randomly selected individuals per population (50 females/50 males), and ADMIXTURE was used to generate *Q*-scores for each individual. Hybrids were identified using the same $Q < 0.70$ threshold applied above.

*Differentiation across the genome*

We examined genetic differentiation across the genome by pairing our data with the *artedi* linkage map constructed by Blumstein (2019). Catalog IDs were identical between the current study and Blumstein (2019) therefore no alignment step was needed to compare loci. To identify putative genomic islands of

312 divergence that were highly differentiated from the rest of the genome, we used a Gaussian kernel
313 smoothing technique (Gagnaire, Pavey, et al., 2013; Hohenlohe et al., 2010; Larson et al., 2017) that
314 incorporated locus-specific differentiation and genomic position on the linkage map. A window size of 5
315 cM and a stepwise shift of 1 cM was used for this analysis, and values of genetic differentiation were
316 weighted according to their window position as described by Gagnaire et al. (2013). Highly differentiated
317 windows were identified by randomly sampling N loci from the genome (where N was the number of loci
318 in the window) and comparing the average differentiation of those loci to the average differentiation of
319 the loci in the window. This sampling routine was conducted 1,000 times for each window. If a window
320 exceeded the 90$^{th}$ percentile of the sampling distribution, the number of bootstrap replicates was increased
321 to 10,000. Contiguous windows that contained at least two loci and exceed the 99$^{th}$ percentile of the
322 distribution after 10,000 bootstrap replicates were classified as putative islands of divergence. We
323 investigated genomic differentiation using four locus-specific metrics: overall $F_{ST}$ and pairwise $F_{ST}$
324 between each of the three putative forms with n>10 (ART–HOY, ART–KIY, and HOY–KIY). We then
325 plotted overall $F_{ST}$ across the genome and constructed a bubble plot to visualize the number of significant
326 windows on each LG for each comparison. We found over 50 significant windows for each comparison
327 (see results). Conducting an in-depth investigation of all significant windows was not feasible, therefore
328 we isolated in-depth analysis to one LG (Cart21) that contained the most significant windows in the
329 dataset.

330 To investigate this highly differentiated LG, we aligned consensus sequences for all loci from Cart21 to
331 chromosome Ssa05 in Atlantic salmon (*Salmo salar*), which is syntenic to Cart21 in *artedi* (Blumstein,
332 2019). Sequence was obtained from genome version ICSASG_v2 (Lien et al., 2016) and alignments were
333 conducted with BLASTN. The best alignment for each locus was retained, and all alignments had e-
334 values < 1e-58. We then visualized the relationship between recombination and physical distance using
335 alignments to Ssa05 and information from the *artedi* linkage map to determine whether this highly
336 differentiated region is characterized by lower recombination. We also obtained annotation information

337 from the Atlantic salmon genome to determine whether genes of interest were co-located with areas of
338 high divergence. Finally, we plotted the allele frequencies of the 10 SNPs on Cart21 with the highest
339 overall $F_{ST}$ to investigate whether these SNPs show consistent patterns of population structure.

340 **Results**

341 *Sequencing and genotyping*

342 A total of 137 individuals were RAD sequenced producing more than 455 million reads and an average of
343 3,346,457 reads per sample. After filtering, 119 individuals with representatives from all five putative
344 cisco forms in Lake Superior were genotyped at 29,068 loci (Table 1). More than half of these loci
345 (n=15,348 loci) were also placed on the linkage map. Since both RAD sequencing and microsatellite
346 amplification were performed on the same *hoyi* and *kiyi* samples, microsatellite genotypes used in our
347 analyses were restricted to the same individuals that successfully genotyped with our RAD loci. An
348 additional 30 *artedi* from Stockton Island were genotyped at the 9 microsatellite loci. Paired-end
349 assemblies for each locus are available from Blumstein (2019).

350 *Genetic differentiation and diversity*

351 PCA showed a sharp contrast in resolution between marker sets with microsatellites producing one large
352 cluster of overlapped forms across the first two principal components and the RAD dataset producing
353 three major clusters primarily composed of ART, HOY, and KIY (Fig. 2). The ART cluster separates
354 from HOY and KIY along the first principal component (PC), and HOY and KIY form discrete clusters
355 along the second PC with three exceptions. Three individuals originally identified as *hoyi* fell within the
356 KIY cluster (see methods and ADMIXTURE results below). Of the rare forms, *nigripinnis* (NIG) loosely
357 grouped in the center of the PCA, with five of the six samples falling out between the ART, HOY, and
358 KIY clusters (Fig. 2). The sixth sample fell within the ART cluster, and like the three *hoyi* in the KIY
359 cluster, possibly represents a misidentified specimen. Unlike the NIG samples, which suggest the
360 possibility for a distinct cluster with the addition of more specimens, the four ZEN samples closely

associated with either the HOY cluster (n=1) or the KIY cluster (n=3). Low representation of both NIG and ZEN in our RAD dataset reduces our ability to draw strong conclusions based on these PCA results, so both groups were unaltered for estimates of diversity, inbreeding and differentiation and dropped for the remaining analyses.

The most supported number of ancestral populations (*K*) estimated using the cross-validation procedure in ADMIXTURE was two (Fig. 3). Examining additional *K*s for significant sub-structuring among forms generated results that corroborated those from the PCA. When *K*=2, the ART cluster splits from HOY, KIY, and ZEN. Individuals in the NIG cluster exhibited mixed ancestry between the two major groups as seen on PC1 of the PCA. When *K*=3, the major genetic ancestries differentiate the ART, HOY, and KIY clusters as seen on PC2. The three putatively misidentified *hoyi* first noted in the PCA all had *Q* estimates of 100% for the KIY cluster and were removed from further analyses. Additional *K*s did not differentiate either the NIG or ZEN cluster but begin to differentiate small subsets of individuals within groups, which was likely statistical noise.

Observed and expected heterozygosity ranged from 0.186-0.246 and 0.242-0.273 (respectively) in the RAD dataset and 0.534-0.629 and 0.607-0.659 (respectively) in the microsatellite dataset (Table 1). Inbreeding coefficients were not substantially different from zero in both datasets. The largest $G_{IS}$ was measured in ZEN from only four samples (0.279) and the rest were between -0.028-0.196. All estimates of pairwise genetic differentiation among forms with n>10 (ART, HOY, KIY) with the RAD dataset were significant (Table 2). The magnitude of genetic differentiation followed similar trends observed in the PCA, with ART being slightly more differentiated from deepwater forms ($F_{ST}$=0.049-0.056). This pattern remained the same with a standardized measure of genetic differentiation in the RAD and microsatellite datasets (Table 3). All pairwise comparisons in both datasets were significant with higher overall values of $G'_{ST}$ generated using microsatellites (0.110-0.122) than SNPs (0.060-0.75). See Tables S1 and S2 for locus-specific summary statistics.

Population self-assignment rate using the microsatellite dataset ranged from 61.9-76.7% with *artedi* exhibiting the highest likelihood of being assigned back to the ART group (Table 1). Assignment rate with the RAD dataset was 100% for the KIY group and 98.3% and 92.5% for the ART and HOY groups (respectively), a result of one putative *artedi* assigning to HOY and two putative *hoyi* individuals assigning to KIY. In the ADMIXTURE analysis, these three individuals exhibited relatively high *Q* estimates for ancestry to the populations to which they were assigned (*artedi*, $Q_{HOY}$ = 67.4%; *hoyi*, $Q_{KIY}$= 64.2% and 56.2%). In the PCA generated from the same data, these individuals were oriented between the main clusters. In the microsatellite dataset with the same *hoyi* and *kiyi* samples, neither of the two potentially misclassified *hoyi* assigned to the HOY group, with one being assigned to ART and one to KIY. Assignment scores are reported in supplementary tables S3-S5.

Estimates of $N_e$ with the microsatellite dataset ranged from 73 in HOY to 659 in KIY (Table 1). Only the estimate for ART produced both upper and lower bound confidence intervals ($N_e$: 91, CI: 42-6,126), whereas the estimates for HOY and KIY produced confidence intervals with an '*infinite'* upper bound. An '*infinite'* upper bound is typically an indication that the data are not powerful enough to produce an accurate estimate of $N_e$ given the sample size, population size, and/or marker resolution (Do et al. 2014; Waples & Do 2010). For the RAD dataset, estimates of $N_e$ were generated with loci that were placed on the linkage map and ranged from 1,701-2,126 with confidence intervals within 10% of these values.

We documented a relatively high level of hybridization in our empirical dataset, with ART-HOY hybrids comprising 8.6% of the total number of sampled *artedi* and *hoyi*, HOY-KIY hybrids comprising 4.3% of sampled *hoyi* and *kiyi* , and ART-KIY hybrids comprising 0% (Fig. 3, Table S6). Additionally, most of these appear to be F1 hybrids with similar contributions from two genetic groups (Fig. 3). The fact that we observed frequent hybridization coupled with relatively high genetic differentiation was puzzling and prompted us to conduct two types of simulations to investigate how hybridization (i.e. migration) can influence genetic differentiation. Simulated hybridization over a 15-generation period resulted in declines in genetic differentiation among populations for all tested levels of migration (Fig. 4). When the

migration rate was set to 5 or 10%, levels of differentiation more than halved after only five generations for all initial levels of $F_{ST}$. A migration rate of 1% resulted in steadily declining genetic differentiation but only began to approach a halved $F_{ST}$ after 15 generations.

Simulations of stepping stone migration between three populations and subsequent ADMIXTURE analysis resulted in an overall pattern very similar to that observed in our empirical data (Figs 3, S1; Table S6). The major goal of these simulations was to determine whether individuals that we observed in our empirical data with *Q*-scores between 0.1-0.2 for alternative forms were advanced hybrid backcrosses or whether these observations were statistical noise. Results from the simulations with 2 generations of migration (G2), where only F1 migrants are possible, demonstrated that these 0.1-0.2 *Q*-scores are likely statistical noise, as they were present in this simulation (Fig S1). Simulations with 5 (G5) and 10 (G10) generations of migration illustrate that advanced backcrosses will likely have *Q*-scores for alternate forms of at least 0.2. These simulations therefore support our hypothesis that most of the hybrids observed in the empirical data are F1 crosses.

*Differentiation across the genome*

Genetic differentiation across the genome was generally high, with many loci displaying overall $F_{ST}$ values > 0.2 (Figs 5, S2-S5). This differentiation was also not localized to a few LGs, as every LG had at least one locus with $F_{ST} > 0.2$. Kernel smoothing analysis to investigate putative islands of divergence revealed 389 genomic windows that displayed significantly elevated differentiation compared to the rest of the genome (Figs 5, S2-S5). The number of significant windows was highest for the overall $F_{ST}$ comparison (115), followed by the ART-HOY comparison (101), ART-KIY comparison (98), and HOY-KIY comparison (75). Significantly differentiated windows were found on all but five LGs (average per LG=10, SD=10, range=0-36). LG Cart21 displayed the highest number of differentiated windows, prompting us to conduct an in-depth investigation of this LG to investigate patterns and potential drivers of divergence (Fig. 6).

Cart21 contained 36 significantly differentiated windows; with the largest number of windows found for the ART-KIY comparison (18), followed by the overall $F_{ST}$ comparison (13), ART-HOY comparison (4), and HOY-KIY comparison (1). We were able to place 351 loci on Cart21, and 12 of these loci displayed overall $F_{ST}$ values > 0.3 (Fig. 6a). The largest cluster of high-$F_{ST}$ loci was found between 0-10 cM on the linkage map. This region appears to be characterized by relatively low recombination, as loci found in the first 10 cM of Cart21 span about 25 megabases of the Atlantic salmon genome (Fig. 6b). Alignments to the Atlantic salmon genome were possible for 151 loci on Cart21, and these alignments revealed that the highest $F_{ST}$ loci were found between positions 15 million and 25 million on Ssa05 (Fig. 6c). Some of these loci were found in genes with functions that include cell signaling and membrane transport. However, there are over 2,000 genes on Ssa05, making it difficult to reach any robust conclusions about the functional significance of our loci. Allele frequencies at the high-$F_{ST}$ were generally the most diverged between ART and the other two forms, KIY in particular (Fig. 6d).

## Discussion

The resolution of genetic structure in recently diverged species complexes has proved challenging with traditional genetics methods, prompting the reevaluation with genomics of a taxonomically uncertain species complex in the Laurentian Great Lakes. Using 29,068 SNPs to examine differentiation and diversity of sympatric coregonines in the Apostle Islands of Lake Superior, we were able to unambiguously assign individuals to the three major forms as well as identify putative F1 hybrids and mis-identified individuals. Despite a century of anthropogenic impacts in the Great Lakes that has seen the extirpation and extinction of historically documented forms in the *C. artedi* species complex, estimates of $N_e$ and diversity in Apostle Island populations do not suggest the three major extant forms are experiencing bottleneck effects. Genetic differentiation among forms was notably high despite the presence of hybrids. However, simulations to explore the impacts of hybridization on differentiation and the interpretation of low levels of statistical noise in ADMIXTURE analyses indicate a high likelihood that hybridization beyond the F1 state is not successful. This is further supported by the discovery of

widespread differentiation between forms across the genome, indicating that much of the divergence observed has been driven by long term reproductive isolation and drift.

*Hypotheses for high genetic differentiation among forms*

Genetic differentiation of the three primary cisco forms in our study (*artedi*, *hoyi*, *kiyi*) was relatively high compared to previous research in cisco using allozymes, mtDNA, microsatellites, AFLPs, and RAD data, which has largely suggested that forms are not frequently diverged within lakes (but see Stott et al., in press; Turgeon, Estoup, & Bernatchez, 1999; Turgeon et al., 2016). For example, Piette-Lauzière et al. (2019) documented neutral $F_{ST}$ values near or below 0.01 between forms in small lakes within Algonquin Provincial Park (Ontario, Canada) using RAD data, which were much lower than the $F_{ST}$ values of ~0.05 observed in our study. Unfortunately, genetic data for cisco in the Great Lakes is relatively sparse, however, a previous study using mtDNA and microsatellites did not find evidence of differentiation among forms (Turgeon & Bernatchez, 2003). Results from the microsatellites genotyped in the current study are similar and indicate that these markers are unable to differentiate species in Lake Superior, even though genetic structure was relatively high according to the RAD data. Interestingly, our estimates of genetic divergence among forms more closely mirror two studies in lake whitefish and European whitefish that used RAD sequencing (Feulner & Seehausen, 2019; Gagnaire, Pavey, et al., 2013) than previous studies in cisco that genotyped mtDNA and microsatellites.

The high genetic divergence among forms observed in cisco is not typical of other fishes in the Laurentian Great Lakes. Both lake trout (*Salvelinus namaycush*) and brook trout (*Salvelinus fontinalis*) display significant life history polymorphisms in the Great Lakes, with lake trout exhibiting morphologically distinct ecotypes related to depth and brook trout exhibiting both fluvial and adfluvial life histories. Two recent studies used RAD sequencing to investigate life history polymorphism in these species, and neither was able to document strong signals of divergence among forms (Elias, McLaughlin, Mackereth, Wilson, & Nichols, 2018; Perreault-Payette et al., 2017). It is possible that divergence within these forms was reduced through introgression mediated by stocking, as these species were stocked

484 heavily, whereas cisco was not (Baillie, Muir, Scribner, Bentzen, & Krueger, 2016; Wilson et al., 2008).
485 However, it is also possible that reproductive isolation among cisco forms is more complete, reducing the
486 potential for introgression to erode divergence among forms.

487 Very little is known about the spawning biology of forms outside of *artedi*, although our observation of
488 relatively frequent F1 hybrids between ART-HOY and HOY-KIY suggests that there is a least some
489 overlap in reproductive timing among the three forms. It also notable that we did not observe any putative
490 hybrids between ART-KIY. These three cisco forms are encountered in different depths in Lake Superior,
491 with *artedi* inhabiting waters < 80 m deep, *hoyi* inhabiting depths between 60-160 m, and *kiyi* inhabiting
492 depths from 80-200 m (Eshenroder et al., 2016). This depth stratification likely explains our observation
493 that *hoyi* hybridizes with *artedi* and *kiyi* but *artedi* and *kiyi* do not hybridize with each other.

494 Our observation that F1 hybrids are relatively common but hybrid backcrosses appear to be uncommon is
495 consistent with substantially reduced fitness and genetic incompatibilities in hybrid backcrosses.
496 Successful hybridization can homogenize genetic structure within a few generations, as evidenced by our
497 simulations and by a large body of literature in species such as European whitefish and cichlids (reviewed
498 inSeehausen, 2006). However, the fact that we do not observe successful backcrosses suggests that
499 negative interactions between hybrid genomes, such as Dobzhansky-Muller incompatibilities
500 (Dobzhansky, 1936; Muller, 1942), may be present in these backcrosses (Dagilis, Kirkpatrick, & Bolnick,
501 2019; Ellison & Burton, 2008). Multiple lines of evidence for hybrid incompatibilities have been found in
502 a sister taxon of cisco, lake whitefish (*Coregonus clupeaformis*, reviewed in Bernatchez et al., 2010). For
503 example, Rogers and Bernatchez (2006) found that hybrid backcrosses had ~6 times higher mortality than
504 F1 crosses, Renaut, Nolte, and Bernatchez (2009) found disruption of gene expression pattern in
505 backcrosses, and Whiteley, Persaud, Derome, Montgomerie, and Bernatchez (2009) documented reduced
506 sperm performance in backcrosses. Our results combined with those from lake whitefish provide evidence
507 that hybrid backcrosses may experience dramatically reduced fitness in cisco. However, future research is

508 necessary to empirically test this hypothesis and investigate potential mechanisms for genomic
509 incompatibilities in hybrid backcrosses.

510 *Genetic differentiation across the genome*

511 Comparison of patterns of genomic divergence found in our study with previous research suggests that
512 diversification of cisco forms in the Great Lakes is likely polygenic and that these forms have been
513 isolated without gene flow for a relatively long period of time. We identified over 100 significantly
514 differentiated genomic windows in our study, and genetic differentiation among forms was consistently
515 high across the genome. This result is similar to two other investigations of genomic divergence in
516 coregonines (Feulner & Seehausen, 2019; Gagnaire, Pavey, et al., 2013), but differs substantially from
517 genome scans in some other salmonids that have revealed supergenes responsible for substantial
518 phenotypic divergence across large geographic areas and differing genetic backgrounds (e.g. Barson et
519 al., 2015; Pearse et al., 2018; Prince et al., 2017). Specifically, Feulner and Seehausen (2019) investigated
520 genomic divergence in three species of whitefishes (*Coregonus* spp) from two lakes in Switzerland and
521 found high divergence on all chromosomes, and Gagnaire et al. (2013) assessed divergence between
522 species pairs of lake whitefish (*C. clupeaformis*) across five lakes in the St. John River Basin (Maine,
523 USA and Québec, Canada) and also found consistently high differentiation across the genome. By
524 comparison, Barson et al. (2015) found a single genomic region that is largely responsible for age-at-
525 maturity in Atlantic salmon (*Salmo salar*), Prince et al. (2017) discovered a small genomic region that is
526 highly associated with run timing in steelhead (*Oncorhynchus mykiss*) and Chinook salmon
527 (*Oncorhynchus tshawytscha*), and Pearse et al. (2018) documented a large inversion associated with life
528 history polymorphism (anadromy vs residency) in steelhead. Taken together, these results suggest
529 divergence of coregonines in both Europe and North America is likely polygenic. This conclusion is
530 supported by a large body of research in lake whitefish demonstrating that traits that differ among species
531 (e.g. body size, growth rate) are controlled by many quantitative trait loci (QTL) (Gagnaire, Normandeau,
532 et al., 2013; Laporte et al., 2015; Rougeux, Gagnaire, Praebel, Seehausen, & Bernatchez, 2019).

Additionally, signatures of parallel evolution at the genomic level are relatively rare in coregonines (Feulner & Seehausen, 2019; Gagnaire, Pavey, et al., 2013), providing evidence that adaptive divergence in this taxon is likely controlled by many genes of small effect that generally differ across speciation events.

Even when speciation is controlled by many genes of small effect, patterns of genetic differentiation may not be homogenous across the genome (Nosil & Feder, 2012; Nosil, Funk, & Ortiz-Barrientos, 2009). Heterogenous genomic divergence is hypothesized to be especially common during sympatric speciation (i.e. divergence with gene flow), as selectively advantageous loci that are clustered together are protected from between population recombination (Via, 2012; Via & West, 2008). Prolific islands of divergence have been identified in stickleback (*Gasterosteus aculeatus*) forms that diverged within the last 150 years (Marques et al., 2016) and in species with high gene flow, such as stick insects (*Timema cristinae*, Soria-Carrasco et al., 2014) and *Heliconius* butterflies (Nadeau et al., 2012). Often these islands are found in areas of low recombination (Burri et al., 2015; Samuk et al., 2017), as we observed in LG Cart21. However, as populations drift apart, high levels of divergence accumulate across the genome, making it difficult to differentiate genomic islands involved in the original divergence from accumulated genetic drift in populations that have been isolated for a long time period (Via, 2012). This pattern was observed in sockeye salmon (*Oncorhynchus nerka*), where an island of divergence that is highly visible across multiple drainages in Alaska was partially obscured in the Copper River, which has experienced lower gene flow and therefore much more genetic drift than the other systems (Larson et al., 2019). The high and relatively homogenous patterns of genetic differentiation observed in our study suggest that genomic islands involved in diversification may be obscured in a similar fashion.

*Conservation implications*

We documented high genetic differentiation among the three major cisco forms in Lake Superior and, based on this information, we suggest that separate conservation units could be constructed for each form. This strategy differs from the current conservation paradigm for cisco, which recommends that the entire

558 *C. artedi* species complex be considered *C. artedi* sensu lato (translation, in the broad sense) and that
559 units of conservation should be designed to preserve environments that have facilitated the evolution of
560 different forms (i.e. lakes) rather than on forms at a larger scale (Turgeon & Bernatchez, 2003; Turgeon et
561 al., 2016). These recommendations were informed by the best available data, which up to this point, have
562 been generated with commonly employed markers for the detection of taxonomic and conservation units -
563 AFLPs, mtDNA, and microsatellites (Reed et al., 1998; Turgeon & Bernatchez, 2003; Turgeon et al.,
564 2016) - none of which have been able to consistently resolve different forms within the Great Lakes (but
565 see Stott et al., in press for a single lake example). Our results suggest that "last generation" markers may
566 be insufficient for capturing differentiation in evolutionarily young species, such as cisco, and highlight
567 the utility of genomic data for designating conservation units in these species. However, it is important to
568 note that we only surveyed animals from one lake, and a larger genotyping effort across the Great Lakes
569 is necessary to accurately inform conservation units for Great Lakes ciscoes. Additionally, our ability to
570 draw conclusions regarding the relationships of the rare forms *nigripinnis* and *zenithicus* to the three
571 major forms was limited by low sample size, therefore, genotyping additional contemporary and/or
572 historic specimens may help resolve the placement these forms.

573 The potential of genomic data to revolutionize construction of conservation units has been frequently
574 discussed (reviewed in Allendorf et al., 2010; Funk et al., 2012), and many studies have found that
575 genomic data provides increased resolution for delineating population structure compared to last
576 generation markers, such as microsatellites (Hodel et al., 2017; Vendrami et al., 2017; Wagner et al.,
577 2013). However, conservation units that were constructed with these last generation markers have
578 generally proven to be robust and are usually only updated slightly, if at all, based on genomic data (e.g.
579 Hecht, Matala, Hess, & Narum, 2015; Larson et al., 2014; Moore et al., 2014). Our findings do not follow
580 this pattern and suggest that current conservation units for cisco may not accurately reflect the diversity of
581 this species. This finding has larger implications for constructing conservation units. Specifically, our data
582 suggest that conservation units constructed based on data from last generation markers that found either

583 low or no significant structure among groups of animals with different phenotypes could potentially be re-
584 evaluated with genomic tools to ensure within-species diversity is being adequately conserved.

585 *Conclusions and future directions*

586 Our study provides the first evidence that cisco forms within the Great Lakes are genetically
587 differentiated. We documented high genetic differentiation among the three major forms in Lake
588 Superior, and highly differentiated markers were distributed across the genome, with islands of
589 divergence found on nearly every linkage group. Additionally, we identified putative F1 hybrids but no
590 hybrid backcrosses, suggesting that fitness breakdown of backcrosses may aid in maintaining
591 differentiation of these forms. The results of this study provide the foundation for a new understanding of
592 the ecology and evolution of the *C. artedi* species complex within the Great Lakes. The ability to
593 differentiate forms with genomics provides researchers with a powerful tool for ground truthing
594 morphological phenotypes and identifying cisco species at any life stage. In particular, the ability to
595 identify larval ciscoes will allow researchers to estimate recruitment, which is vital for management and
596 conservation, and will also significantly improve our understanding of early life history characteristics
597 and reproductive dynamics in this species. Finally, our results suggest that some management units
598 created using last generation markers may not adequately reflect species diversity and could be re-
599 evaluated with genomic data.

## Tables

**Table 1.** Sample statistics, diversity, and effective population size estimates. **N** is the number of individuals successfully genotyped, **A** is the percentage of total sampled alleles found in each group, $H_o/H_e$ are observed/expected heterozygosity, $G_{IS}$ is the inbreeding coefficient, **Assignment** is the percentage of individuals that were correctly assigned to their population of origin in a leave-one-out test, and $N_e$ is effective population size calculated using the LDNE method and reported with 95% confidence intervals.

| Method | Group | Code | N | A | $H_o$ | $H_e$ | $G_{IS}$ | Assignment | $N_e$ (CIs) |
|---|---|---|---|---|---|---|---|---|---|
| RAD | *C. artedi* | ART | 60 | 98.5% | 0.246 | 0.253 | 0.028 | 98.3%** | 1,834 (1,769-1,906) |
| | *hoyi.* | HOY | 21* | 94.8% | 0.219 | 0.273 | 0.196 | 92.5%** | 1,701 (1,487-1,985) |
| | *kiyi* | KIY | 25 | 93.4% | 0.217 | 0.242 | 0.103 | 100% | 2,126 (1,846-2,506) |
| | *nigripinnis* | NIG | 6 | 83.1% | 0.213 | 0.261 | 0.182 | - | - |
| | *zenithicus* | ZEN | 4 | 76.5% | 0.186 | 0.259 | 0.279 | - | - |
| Microsatellites | *C. artedi* | ART | 30 | 70.0% | 0.534 | 0.607 | 0.121 | 76.7% | 91 (42-6,126) |
| | *hoyi* | HOY | 21* | 65.0% | 0.629 | 0.659 | 0.046 | 61.9% | 73 (33-Infinite) |
| | *kiyi* | KIY | 25 | 67.0% | 0.572 | 0.631 | 0.093 | 72.0% | 649 (74-Infinite) |

*Three genotyped *hoyi* samples are suspected to be misidentified *kiyi* from the RAD-based PCA and ADMIXTURE analysis and were removed to prevent bias in estimates of diversity and $N_e$.
**The one *artedi* that assigned to HOY and the two *hoyi* that assigned to KIY appear to be hybrids. Outside of these putative hybrids, assignment to the ART and HOY groups was 100%.

**Table 2**. Pairwise differentiation among putative forms calculated with the RAD dataset. $F_{ST}$ values on the lower diagonal. Significant values are in bold.

| Group | ART | HOY | KIY | NIG | ZEN |
|---|---|---|---|---|---|
| **ART** | - | | | | |
| **HOY** | **0.049** | - | | | |
| **KIY** | **0.056** | **0.045** | - | | |
| **NIG** | 0.019 | **0.013** | **0.034** | - | |
| **ZEN** | **0.061** | **0.017** | 0.020 | 0.014 | - |

**Table 3**: Standardized pairwise genetic differentiation, $G'_{ST}$, (Hedrick, 2005) for SNP and microsatellite datasets. Values measured using the SNP dataset are in the lower diagonal, and values measured using the microsatellite dataset are in the upper diagonal.

| Group | ART | HOY | KIY |
|---|---|---|---|
| **ART** | - | 0.122 | 0.113 |
| **HOY** | 0.064 | - | 0.110 |
| **KIY** | 0.075 | 0.060 | - |

## Figures

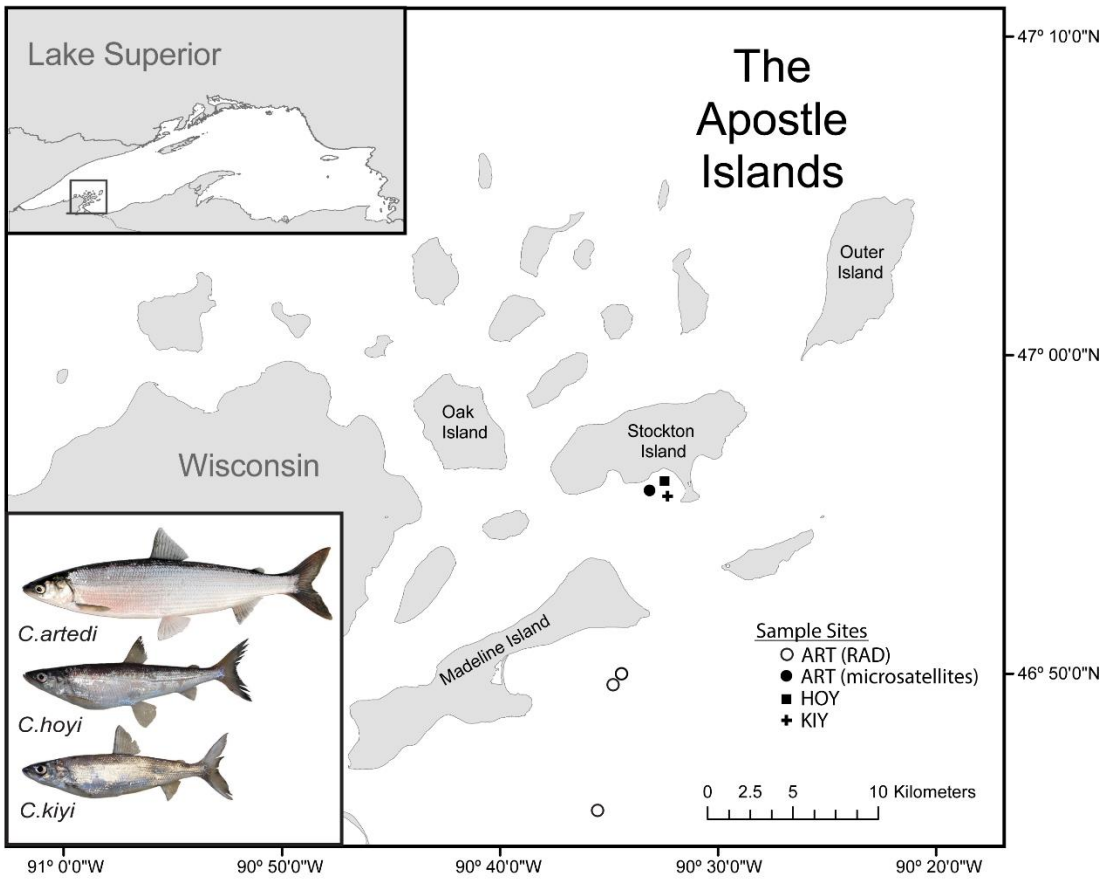



**Fig. 1**. Map of sample sites in the Apostle Islands. INSET (upper right): Location of the Apostle Islands, gray box, within Lake Superior. INSET (lower right): Photos of the three most common cisco forms found in Lake Superior from Eshenroder et al. (2016, used with permission from the author).

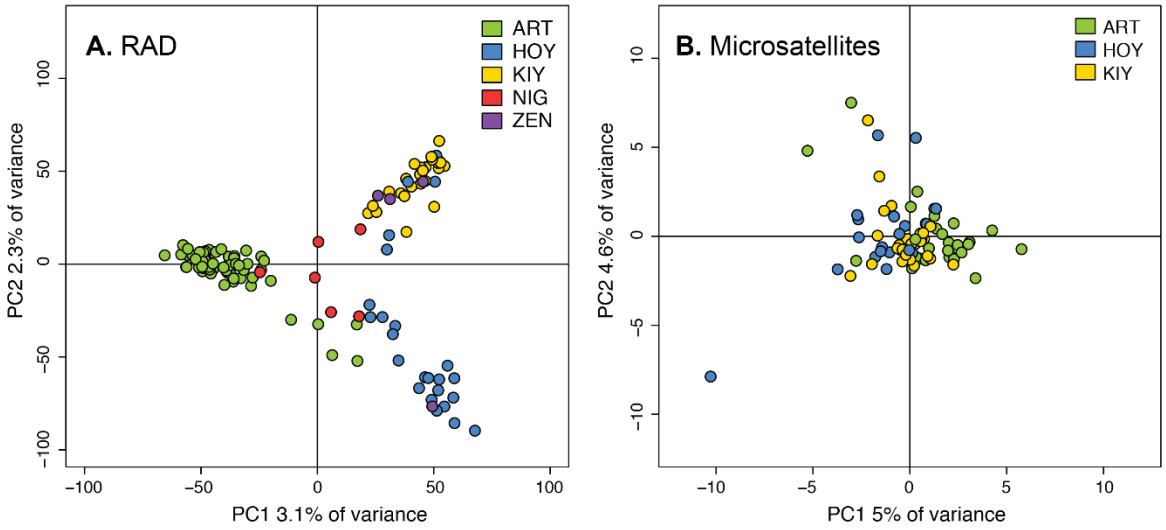
628

629

630

631

632

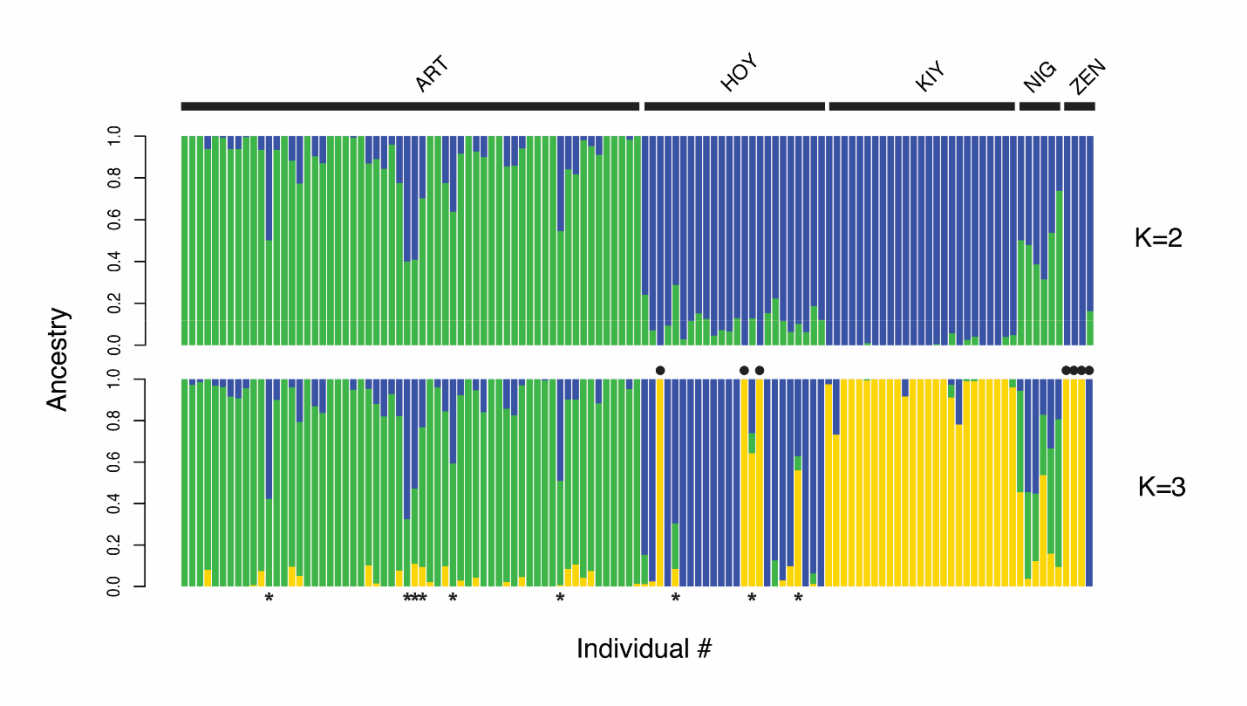



633

634 **Fig. 2**. Principal components analysis with RAD (A) and microsatellite (B) data. The percentage of

635 variance explained by each principal component (PC) is labeled on the x- and y- axes.

636




637

638 **Fig. 3**. Genetic lineages in Apostle Island ciscoes estimated with ADMIXTURE. Each vertical bar
639 represents a single individual and is colored by the proportion of ancestry (*Q*) assigned to each genetic
640 lineage (*K*). In the *K*=3 plot (bottom), filled circles represent putatively misidentified individuals within
641 forms (with *Q*-scores of 100% to other forms) and asterisks represent putative hybrids based on a *Q*-score
642 threshold of less than 70%.

643

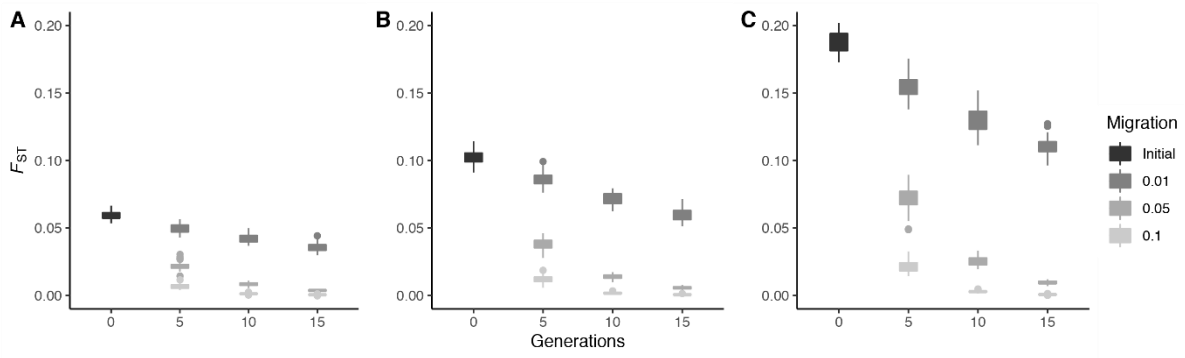



**Fig. 4**. Simulated impacts of differing rates of hybridization under three different levels of preliminary differentiation. Each EASYPOP simulation was run using 1000 biallelic loci in two randomly mating populations each comprised of 1000 females and 1000 males with 50 replicates of each unique set of parameters for three starting levels of differentiation: A) a migration rate of 0.001 over 1000 generations resulting in an $F_{ST}$≈0.05 followed by 5, 10, or 15 generations of hybridization at 1, 5, and 10% per generation, B) a migration rate of 0.0005 over 1000 generations resulting in an $F_{ST}$≈0.10 followed by 5, 10, or 15 generations of hybridization at 1, 5, and 10% per generation, and C) a migration rate of 0.0001 over 1000 generations resulting in an $F_{ST}$≈0.20 followed by 5, 10, or 15 generations of hybridization at 1, 5, and 10% per generation.

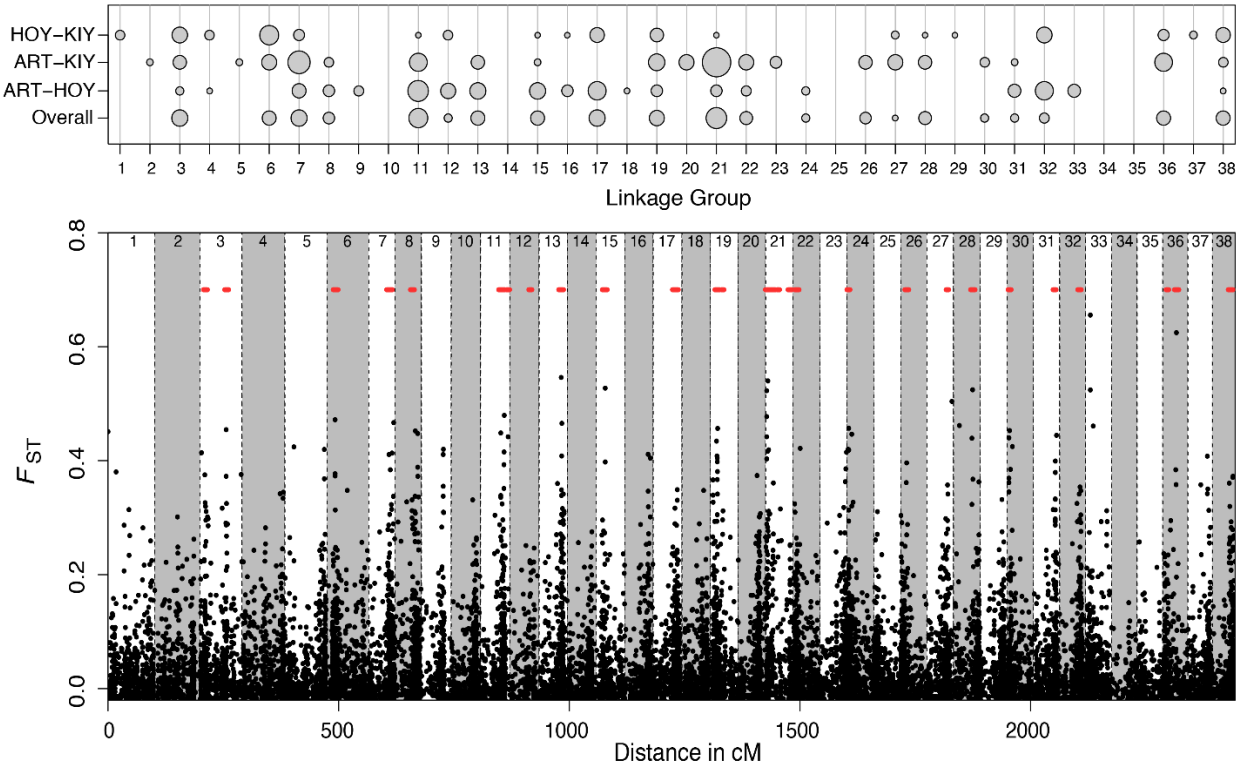



**Fig. 5**. Genetic differentiation across the genome visualized with a bubble plot (top) and plot with the overall $F_{ST}$ of each marker (bottom). The size of each bubble in the bubble plot represents the number of genomic windows that were significantly differentiated from the rest of the genome according to kernel smoothing analysis for each form comparison. The 'overall' designation is overall $F_{ST}$ across the dataset. Each black dot in the graph of differentiation across the genome represents a marker, and red lines denote significantly differentiated windows. Linkage groups are separated by dashed lines. Form abbreviations are in Table 1. See Figs. S2-S5 for visualizations of genetic differentiation for each chromosome and form comparison.

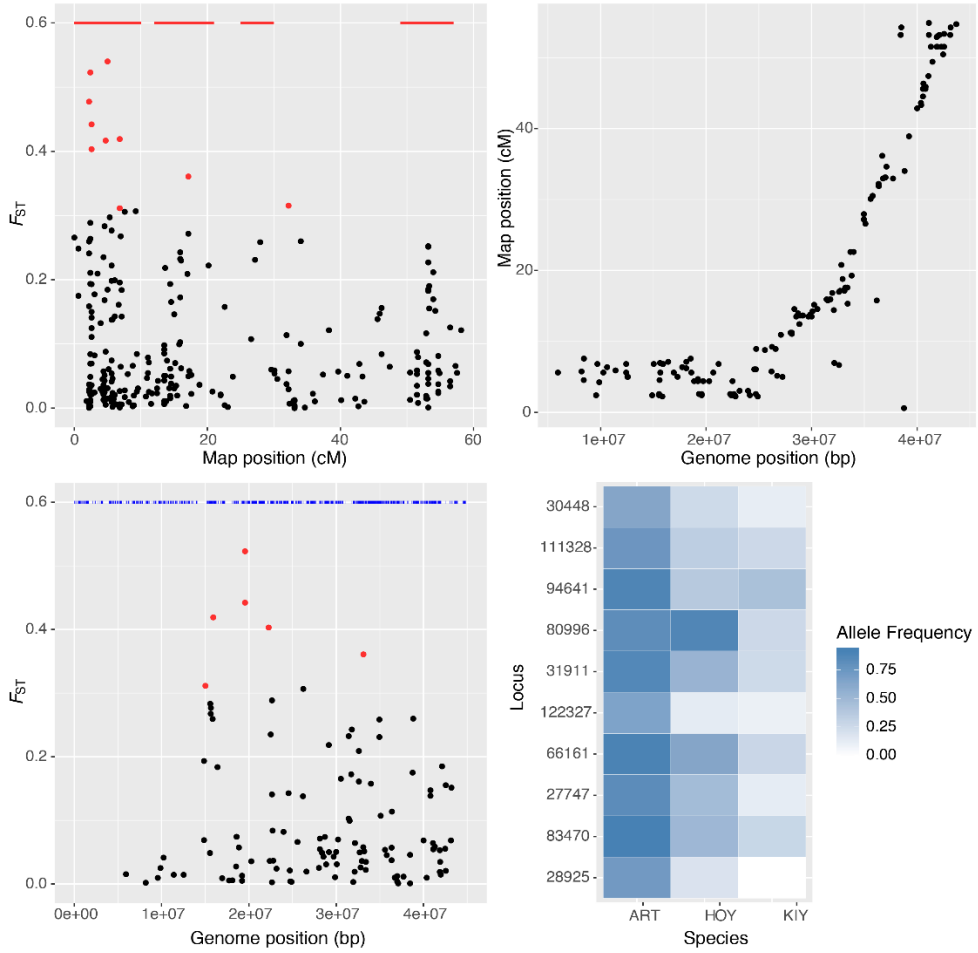



**Fig. 6**. Investigation of genetic differentiation on linkage group Cart21, the linkage group with the most significantly differentiated windows. (a) Genetic differentiation (overall $F_{ST}$) at 351 loci that were placed on Cart21 in the cisco linkage map. The top ten loci with the highest $F_{ST}$ are colored red. Red lines denote significantly differentiated windows. (b) Recombination distance on cisco linkage group Cart21 (y-axis) versus physical distance on Atlantic salmon chromosome Ssa05 (x-axis). (c) Genetic differentiation (overall $F_{ST}$) at 152 loci from Cart21 that successfully aligned to Ssa05, the syntenic chromosome in Atlantic salmon. Six of the top ten loci from panel a aligned to Ssa05 and are colored red. Blue lines indicate the position of genes on Ssa05. (d) Allele frequencies of the loci with the highest $F_{ST}$ from Cart21. Loci are ordered from highest $F_{ST}$ (bottom) to lowest. Form abbreviations are found in Table 1.

675 **Supplementary data:**

676 **Table S1**. Summary statistics for 26,789 loci genotyped with RAD sequencing. The "LG" and "cM"
677 columns denote map location, columns ending in "AF" denote population allele frequencies, "$H_O$" is
678 observed heterozygosity, $H_E$ is expected heterozygosity, the "Sequence P1 column" is the sequence from
679 the P1 read for each RAD tag, and the "Sequence PE" is the sequence obtained from paired-end
680 assemblies. Form abbreviations are identical to those in Table 1.

681 **Table S2**. Summary statistics for 9 microsatellite loci included in this study. Abbreviations are total
682 number of alleles per locus (*A*), effective number of alleles per locus (*Neff*), observed heterozygosity
683 ($H_O$), and expected heterozygosity ($H_E$). $F_{ST}$ (Weir and Cockerham, 1984) was estimated in Genepop and
684 all other statistics were estimated in Genodive.

685 **Table S3.** Initial assignment scores from Genodive for all sampled individuals using RAD data.

686 **Table S4.** Assignment scores from Genodive when tests were limited to the three major forms using RAD
687 data.

688 **Table S5.** Assignment scores from Genodive for the three major forms using microsatellite data.

689 **Table S6**. Type, number, and proportion of hybrids observed in empirical data and simulated populations.
690 Data simulated in EASYPOP were comprised of 5,000 biallelic loci from three populations (P1-P3) that
691 experienced stepping-stone migration (m=0.05) for 2, 5, or 10 generations (G2, G5, G10) after 1,000
692 generations of low migration (m=0.001) to approximate an $F_{ST}$ similar to that observed between our three
693 major cisco forms (ART, HOY & KIY, $F_{ST} \approx 0.05$).

694 **Fig S1**. Levels of genetic admixture in three simulated populations after 2, 5, and 10 generations of
695 stepping-stone migration. Simulations were run with random mating of 1,000 females and 1,000 males in
696 each population using 5,000 biallelic loci, and preliminary conditions that produced a similar level of
697 differentiation observed in our RAD dataset among forms ($F_{ST} \approx 0.05$; 1,000 generations with an island
698 migration rate of 0.001). Each ADMIXTURE plot represents a random subset of 50 males and 50 females
699 from each population.

700 **Fig S2**. Overall $F_{ST}$ for all loci that could be placed on the linkage map. Each dot is a marker and red lines
701 indicate genomic windows that were significantly differentiated from the rest of the genome according to
702 kernel smoothing analysis.

703 **Fig S3**. Pairwise-$F_{ST}$ for the ART-HOY comparison for all loci that could be placed on the linkage map.
704 Each dot is a marker and red lines indicate genomic windows that were significantly differentiated from
705 the rest of the genome according to kernel smoothing analysis. Form abbreviations are in Table 1.

706 **Fig S4**. Pairwise-$F_{ST}$ the ART-KIY comparison for all loci that could be placed on the linkage map. Each
707 dot is a marker and red lines indicate genomic windows that were significantly differentiated from the rest
708 of the genome according to kernel smoothing analysis. Form abbreviations are in Table 1.

709 **Fig S5**. Pairwise-$F_{ST}$ the HOY-KIY comparison for all loci that could be placed on the linkage map. Each
710 dot is a marker and red lines indicate genomic windows that were significantly differentiated from the rest
711 of the genome according to kernel smoothing analysis. Form abbreviations are in Table 1.

712

713

## Data Archiving Statement

Upon acceptance, demultiplexed sequence data used in this research along with corresponding metadata will be archived in the NCBI sequence read archive using the publicly accessible Genomic Observatories Metadatabase (GeOMe, http://www.geome-db.org/) and microsatellite genotypes will be archived on DRYAD.

## Literature Cited

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research, 19*(9), 1655-1664. doi:10.1101/gr.094052.109

Ali, O. A., O'Rourke, S. M., Amish, S. J., Meek, M. H., Luikart, G., Jeffres, C., & Miller, M. R. (2016). RAD Capture (Rapture): Flexible and Efficient Sequence-Based Genotyping. *Genetics, 202*(2), 389-400. doi:10.1534/genetics.115.183665

Allendorf, F., & Luikart, G. (2007). *Conservation and the genetics of populations*: Blackwell Publishing Ltd.

Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature reviews genetics, 11*(10), 697.

Als, T. D., Hansen, M. M., Maes, G. E., Castonguay, M., Riemann, L., Aarestrup, K. I. M., . . . Bernatchez, L. (2011). All roads lead to home: panmixia of European eel in the Sargasso Sea. *Molecular Ecology, 20*(7), 1333-1346. doi:10.1111/j.1365-294X.2011.05011.x

Angers, B., Bernatchez, L., Angers, A., & Desgroseillers, L. (1995). Specific microsatellite loci for brook charr reveal strong population subdivision on a microgeographic scale. *Journal of Fish Biology, 47*(sA), 177-185. doi:10.1111/j.1095-8649.1995.tb06054.x

Baillie, S. M., Muir, A. M., Scribner, K., Bentzen, P., & Krueger, C. C. (2016). Loss of genetic diversity and reduction of genetic distance among lake trout Salvelinus namaycush ecomorphs, Lake Superior 1959 to 2013. *Journal of Great Lakes Research, 42*(2), 204-216. doi:https://doi.org/10.1016/j.jglr.2016.02.001

Balloux, F. (2001). EASYPOP (Version 1.7): A Computer Program for Population Genetics Simulations. *Journal of Heredity, 92*(3), 301-302. doi:10.1093/jhered/92.3.301

Barson, N. J., Aykanat, T., Hindar, K., Baranski, M., Bolstad, G. H., Fiske, P., . . . Primmer, C. R. (2015). Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature, 528*, 405. doi:10.1038/nature16062

Bernardi, G., & Lape, J. (2005). Tempo and mode of speciation in the Baja California disjunct fish species Anisotremus davidsonii. *Molecular Ecology, 14*(13), 4085-4096. doi:10.1111/j.1365-294X.2005.02729.x

Bernatchez, L. (1996). *Réseau de suivi environnemental du complexe la Grande. Caractérisation génétique des formes naines et normales de grand corégone du réservoir Caniapiscau et du lac Sérigny à l'aide de marqueurs microsatellites.* Retrieved from

Bernatchez, L., Colombani, F., & Dodson, J. (1991). Phylogenetic relationships among the subfamily Coregoninae as revealed by mitochondrial DNA restriction analysis. *Journal of Fish Biology, 39*, 283-290.

Bernatchez, L., & Dodson, J. J. (1994). Phylogenetic relationships among Palearctic and Nearctic whitefish (Coregonus sp.) populations as revealed by mitochondrial DNA variation. *Canadian Journal of Fisheries and Aquatic Sciences, 51*(S1), 240-251.

Bernatchez, L., Renaut, S., Whiteley, A. R., Derome, N., Jeukens, J., Landry, L., . . . St-Cyr, J. (2010). On the origin of species: insights from the ecological genomics of lake whitefish. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 365*(1547), 1783-1800. doi:10.1098/rstb.2009.0274

761 Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K. L., Meier, R., Winker, K., . . . Das, I. (2007). Cryptic
762 species as a window on diversity and conservation. *Trends in Ecology & Evolution, 22*(3), 148-
763 155. doi:https://doi.org/10.1016/j.tree.2006.11.004
764 Blumstein, D. (2019). *The first haploid linkage map in a coregonid (Coregonus artedi) improves*
765 *knowledge of chromosomal evolution and rediploidization across salmonids.* (Masters),
766 University of Wisconsin Stevens Point,
767 Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., . . . Ellegren, H. (2015).
768 Linked selection and recombination rate variation drive the evolution of the genomic landscape of
769 differentiation across the speciation continuum of Ficedula flycatchers. *Genome Research,*
770 *25*(11), 1656-1665.
771 Catchen, J. M., Amores, A., Hohenlohe, P. A., Cresko, W. A., & Postlethwait, J. H. (2011). Stacks:
772 Building and Genotyping Loci De Novo From Short-Read Sequences. *G3 (Bethesda), 1*(3), 171-
773 182. doi:10.1534/g3.111.000240
774 Catchen, J. M., Hohenlohe, P., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis
775 tool set for population genomics. *Mol Ecol, 22*(11), 3124-3140. doi:10.1111/mec.12354
776 Coates, D. J., Byrne, M., & Moritz, C. (2018). Genetic diversity and conservation units: Dealing with the
777 species-population continuum in the age of genomics. *Frontiers in Ecology and Evolution, 6*,
778 165.
779 Commission, G. L. F., & Christie, W. (1973). *Review of the Changes in the Fish species composition of*
780 *Lake Ontario*.
781 Conrad, D. F., Keebler, J. E. M., DePristo, M. A., Lindsay, S. J., Zhang, Y., Casals, F., . . . Genomes, P.
782 (2011). Variation in genome-wide mutation rates within and between human families. *Nature*
783 *genetics, 43*(7), 712-714. doi:10.1038/ng.862
784 Dagilis, A. J., Kirkpatrick, M., & Bolnick, D. I. (2019). The evolution of hybrid fitness during speciation.
785 *PLOS Genetics, 15*(5), e1008125. doi:10.1371/journal.pgen.1008125
786 Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., . . . Genomes Project
787 Analysis, G. (2011). The variant call format and VCFtools. *Bioinformatics, 27*(15), 2156-2158.
788 doi:10.1093/bioinformatics/btr330
789 Do, C., Waples, R. S., Peel, D., Macbeth, G., Tillett, B. J., & Ovenden, J. R. (2014). NeEstimator v2: re-
790 implementation of software for the estimation of contemporary effective population size (Ne)
791 from genetic data. *Molecular ecology resources, 14*(1), 209-214.
792 Dobzhansky, T. (1936). Studies on Hybrid Sterility. II. Localization of Sterility Factors in Drosophila
793 Pseudoobscura Hybrids. *Genetics, 21*(2), 113-135.
794 Elias, A., McLaughlin, R., Mackereth, R., Wilson, C., & Nichols, K. M. (2018). Population structure and
795 genomic variation of ecological life history diversity in wild-caught Lake Superior brook trout,
796 Salvelinus fontinalis. *Journal of Great Lakes Research, 44*(6), 1373-1382.
797 doi:https://doi.org/10.1016/j.jglr.2018.08.006
798 Ellison, C. K., & Burton, R. S. (2008). Interpopulation hybrid breakdown maps to mitochondrial genome.
799 *Evolution, 62*(3), 631-638. doi:10.1111/j.1558-5646.2007.00305.x
800 Eshenroder, R. L., Vecsei, P., Gorman, O. T., Yule, D., Pratt, T. C., Mandrak, N. E., . . . Muir, A. M.
801 (2016). *Ciscoes (Coregonus, subgenus Leucichthys) of the Laurentian Great Lakes and Lake*
802 *Nipigon*. Retrieved from http://pubs.er.usgs.gov/publication/70179454
803 Estoup, A., Presa, P., Krieg, F., Vaiman, D., & Guyomard, R. (1993). (CT)n and (GT)n microsatellites: a
804 new class of genetic markers for Salmo trutta L. (brown trout). *Heredity, 71*(5), 488-496.
805 doi:10.1038/hdy.1993.167
806 Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform
807 population genetics analyses under Linux and Windows. *Molecular ecology resources, 10*(3),
808 564-567. doi:10.1111/j.1755-0998.2010.02847.x
809 Feulner, P. G. D., & Seehausen, O. (2019). Genomic insights into the vulnerability of sympatric whitefish
810 species flocks. *Mol Ecol, 28*(3), 615-629. doi:10.1111/mec.14977

811 Funk, W. C., McKay, J. K., Hohenlohe, P. A., & Allendorf, F. W. (2012). Harnessing genomics for
812 delineating conservation units. *Trends in Ecology & Evolution, 27*(9), 489-496.
813 Gagnaire, P.-A., Normandeau, E., Pavey, S. A., & Bernatchez, L. (2013). Mapping phenotypic,
814 expression and transmission ratio distortion QTL using RAD markers in the Lake Whitefish
815 (Coregonus clupeaformis). *Molecular Ecology, 22*(11), 3036-3048. doi:10.1111/mec.12127
816 Gagnaire, P.-A., Pavey, S. A., Normandeau, E., & Bernatchez, L. (2013). The genetic architecture of
817 reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed
818 by RAD sequencing. *Evolution, 67*(9), 2483-2497. doi:10.1111/evo.12075
819 Garside, E., & Christie, W. (1962). Experimental hybridization among three coregonine fishes.
820 *Transactions of the American Fisheries Society, 91*(2), 196-200.
821 Hecht, B. C., Matala, A. P., Hess, J. E., & Narum, S. R. (2015). Environmental adaptation in Chinook
822 salmon (Oncorhynchus tshawytscha) throughout their North American range. *Molecular Ecology,*
823 *24*(22), 5573-5595. doi:10.1111/mec.13409
824 Hedrick, P. W. (2005). A standardized genetic differentiation measure. *Evolution, 59*(8), 1633-1638.
825 doi:10.1111/j.0014-3820.2005.tb01814.x
826 Hey, J., Waples, R. S., Arnold, M. L., Butlin, R. K., & Harrison, R. G. (2003). Understanding and
827 confronting species uncertainty in biology and conservation. *Trends in Ecology & Evolution,*
828 *18*(11), 597-603. doi:https://doi.org/10.1016/j.tree.2003.08.014
829 Hill, W. G. (1981). Estimation of effective population size from data on linkage disequilibrium. *Genetical*
830 *Research, 38*, 7. doi:10.1017/s0016672300020553
831 Himberg, K. (1970). A systematic and zoogeographic study of some North European Coregonids. *Biology*
832 *of coregonid fishes*, 219-250.
833 Hodel, R. G. J., Chen, S., Payton, A. C., McDaniel, S. F., Soltis, P., & Soltis, D. E. (2017). Adding loci
834 improves phylogeographic resolution in red mangroves despite increased missing data:
835 comparing microsatellites and RAD-Seq and investigating loci filtering. *Scientific Reports, 7*(1),
836 17598. doi:10.1038/s41598-017-16810-7
837 Hoey, J. A., & Pinsky, M. L. (2018). Genomic signatures of environmental selection despite near-
838 panmixia in summer flounder. *Evolutionary Applications, 11*(9), 1732-1747.
839 doi:10.1111/eva.12676
840 Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., & Cresko, W. A. (2010).
841 Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD
842 Tags. *PLOS Genetics, 6*(2), e1000862. doi:10.1371/journal.pgen.1000862
843 Hudson, A. G. (2011). *Population Genomics and Ecology of Parallel Adaptive Radiations: The Alpine*
844 *Lake Whitefish*. Verlag nicht ermittelbar,
845 Hudson, A. G., Lundsgaard-Hansen, B., Lucek, K., Vonlanthen, P., & Seehausen, O. (2017). Managing
846 cryptic biodiversity: Fine-scale intralacustrine speciation along a benthic gradient in Alpine
847 whitefish (Coregonus spp.). *Evolutionary Applications, 10*(3), 251-266.
848 Huitfeldt-Kaas, H. (1918). Ferskvandsfiskenes utbredelse og innvandring i Norge medet tillceg om
849 krebsen. Centraltrykkeriet, Kristiania. In: Norwegian.
850 Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers.
851 *Bioinformatics, 24*(11), 1403-1405. doi:10.1093/bioinformatics/btn129
852 Kahilainen, K. K., Østbye, K., Harrod, C., Shikano, T., Malinen, T., & Merilä, J. (2011). Species
853 introduction promotes hybridization and introgression in Coregonus: is there sign of selection
854 against hybrids? *Molecular Ecology, 20*(18), 3838-3855.
855 Kapfer, J. M., Sloss, B. L., Schuurman, G. W., Paloski, R. A., & Lorch, J. M. (2013). Evidence of
856 Hybridization between Common Gartersnakes (Thamnophis sirtalis) and Butler's Gartersnakes
857 (Thamnophis butleri) in Wisconsin, USA. *Journal of Herpetology, 47*(3), 400-405.
858 doi:10.1670/12-057
859 Keenan, K., McGinnity, P., Cross, T. F., Crozier, W. W., & Prodöhl, P. A. (2013). diveRsity: An R
860 package for the estimation and exploration of population genetics parameters and their associated
861 errors. *Methods in Ecology and Evolution, 4*(8), 782-788. doi:10.1111/2041-210X.12067

862 Kimura, M., & Weiss, G. H. (1964). The Stepping Stone Model of Population Structure and the Decrease
863 of Genetic Correlation with Distance. *Genetics, 49*(4), 561-576.
864 Koelz, W. N. (1926). Fishing industry of the Great Lakes.
865 Koelz, W. N. (1929). Coregonid fishes of the Great Lakes. *Bulletin of the United States Bureau of*
866 *Fisheries, 43*(2), 297-643.
867 Laporte, M., Rogers, S. M., Dion-Côté, A.-M., Normandeau, E., Gagnaire, P.-A., Dalziel, A. C., . . .
868 Bernatchez, L. (2015). RAD-QTL mapping reveals both genome-level parallelism and different
869 genetic architecture underlying the evolution of body shape in lake whitefish (Coregonus
870 clupeaformis) species pairs. *G3: Genes, Genomes, Genetics, 5*(7), 1481-1491.
871 Larson, W. A., Dann, T. H., Limborg, M. T., McKinney, G. J., Seeb, J. E., & Seeb, L. W. (2019). Parallel
872 signatures of selection at genomic islands of divergence and the major histocompatibility
873 complex in ecotypes of sockeye salmon across Alaska. *Molecular Ecology, 28*(9), 2254-2271.
874 doi:10.1111/mec.15082
875 Larson, W. A., Limborg, M. T., McKinney, G. J., Schindler, D. E., Seeb, J. E., & Seeb, L. W. (2017).
876 Genomic islands of divergence linked to ecotypic variation in sockeye salmon. *Molecular*
877 *Ecology, 26*(2), 554-570. doi:10.1111/mec.13933
878 Larson, W. A., Seeb, L. W., Everett, M. V., Waples, R. K., Templin, W. D., & Seeb, J. E. (2014).
879 Genotyping by sequencing resolves shallow population structure to inform conservation of
880 Chinook salmon (Oncorhynchus tshawytscha). *Evolutionary Applications, 7*(3), 355-369.
881 doi:10.1111/eva.12128
882 Lien, S., Koop, B. F., Sandve, S. R., Miller, J. R., Kent, M. P., Nome, T., . . . Davidson, W. S. (2016).
883 The Atlantic salmon genome provides insights into rediploidization. *Nature, 533*, 200.
884 doi:10.1038/nature17164
885 Lischer, H. E., & Excoffier, L. (2011). PGDSpider: an automated data conversion tool for connecting
886 population genetics and genomics programs. *Bioinformatics, 28*(2), 298-299.
887 Lu, G., Basley, D., & Bernatchez, L. (2001). Contrasting patterns of mitochondrial DNA and
888 microsatellite introgressive hybridization between lineages of lake whitefish (Coregonus
889 clupeaformis); relevance for speciation. *Molecular Ecology, 10*(4), 965-985.
890 Mann, G., & McCart, P. J. (1981). Comparison of sympatric dwarf and normal populations of least cisco
891 (Coregonus sardinella) inhabiting Trout Lake, Yukon Territory. *Canadian Journal of Fisheries*
892 *and Aquatic Sciences, 38*(2), 240-244.
893 Marie, A., Bernatchez, L., & Garant, D. (2011). Empirical assessment of software efficiency and accuracy
894 to detect introgression under variable stocking scenarios in brook charr (Salvelinus fontinalis).
895 *Conservation Genetics, 12*, 1215-1227. doi:10.1007/s10592-011-0224-y
896 Marques, D. A., Lucek, K., Meier, J. I., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O.
897 (2016). Genomics of Rapid Incipient Speciation in Sympatric Threespine Stickleback. *PLOS*
898 *Genetics, 12*(2), e1005887. doi:10.1371/journal.pgen.1005887
899 McKinney, G. J., Waples, R. K., Seeb, L. W., & Seeb, J. E. (2017). Paralogs are revealed by proportion of
900 heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural
901 populations. *Mol Ecol Resour, 17*(4), 656-669. doi:10.1111/1755-0998.12613
902 Mee, J. A., Bernatchez, L., Reist, J. D., Rogers, S. M., & Taylor, E. B. (2015). Identifying designatable
903 units for intraspecific conservation prioritization: a hierarchical approach applied to the lake
904 whitefish species complex (C oregonus spp.). *Evolutionary Applications, 8*(5), 423-441.
905 Meirmans, P. G., & Hedrick, P. W. (2011). Assessing population structure: FST and related measures.
906 *Molecular ecology resources, 11*(1), 5-18. doi:10.1111/j.1755-0998.2010.02927.x
907 Meirmans, P. G., & Van Tienderen, P. H. (2004). GENOTYPE and GENODIVE: two programs for the
908 analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes, 4*(4), 792-794.
909 Moore, J.-S., Bourret, V., Dionne, M., Bradbury, I., O'Reilly, P., Kent, M., . . . Bernatchez, L. (2014).
910 Conservation genomics of anadromous Atlantic salmon across its North American range: outlier
911 loci identify the same patterns of population structure as neutral loci. *Molecular Ecology, 23*(23),
912 5680-5697. doi:10.1111/mec.12972

913 Muir, A. M., Vecsei, P., Pratt, T. C., Krueger, C. C., Power, M., & Reist, J. D. (2013). Ontogenetic shifts
914 in morphology and resource use of cisco Coregonus artedi. *Journal of Fish Biology, 82*(2), 600-
915 617. doi:10.1111/jfb.12016
916 Muller, H. (1942). *Isolating mechanisms, evolution, and temperature.* Paper presented at the Biol. Symp.
917 Nadeau, N. J., Whibley, A., Jones, R. T., Davey, J. W., Dasmahapatra, K. K., Baxter, S. W., . . . Jiggins,
918 C. D. (2012). Genomic islands of divergence in hybridizing Heliconius butterflies identified by
919 large-scale targeted sequencing. *Philosophical Transactions of the Royal Society B: Biological*
920 *Sciences, 367*(1587), 343-353. doi:10.1098/rstb.2011.0198
921 Nosil, P., & Feder, J. L. (2012). Widespread yet heterogeneous genomic divergence. *Molecular Ecology,*
922 *21*(12), 2829-2832. doi:10.1111/j.1365-294X.2012.05580.x
923 Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic
924 divergence. *Molecular Ecology, 18*(3), 375-402. doi:10.1111/j.1365-294X.2008.03946.x
925 Olsen, M. T., Andersen, L. W., Dietz, R., Teilmann, J., Härkönen, T., & Siegismund, H. R. (2014).
926 Integrating genetic data and population viability analyses for the identification of harbour seal (P
927 hoca vitulina) populations and management units. *Molecular Ecology, 23*(4), 815-831.
928 Østbye, K., Bernatchez, L., Naesje, T., Himberg, K. J., & Hindar, K. (2005). Evolutionary history of the
929 European whitefish Coregonus lavaretus (L.) species complex as inferred from mtDNA
930 phylogeography and gill-raker numbers. *Molecular Ecology, 14*(14), 4371-4387.
931 Paetkau, D., Calvert, W., Stirling, I., & Strobeck, C. (1995). Microsatellite analysis of population
932 structure in Canadian polar bears. *Molecular Ecology, 4*(3), 347-354. doi:10.1111/j.1365-
933 294X.1995.tb00227.x
934 Palm, S., Dannewitz, J., Prestegaard, T., & Wickström, H. (2009). Panmixia in European eel revisited: no
935 genetic difference between maturing adults from southern and northern Europe. *Heredity, 103*,
936 82. doi:10.1038/hdy.2009.51
937 Palsbøll, P. J., Bérubé, M., & Allendorf, F. W. (2007). Identification of management units using
938 population genetic data. *Trends in Ecology & Evolution, 22*(1), 11-16.
939 doi:https://doi.org/10.1016/j.tree.2006.09.003
940 Palsbøll, P. J., Peery, M. Z., & Bérubé, M. (2010). Detecting populations in the 'ambiguous' zone:
941 kinship-based estimation of population structure at low genetic divergence. *Molecular ecology*
942 *resources, 10*(5), 797-805. doi:10.1111/j.1755-0998.2010.02887.x
943 Paris, J. R., Stevens, J. R., Catchen, J. M., & Johnston, S. (2017). Lost in parameter space: a road map for
944 stacks. *Methods in Ecology and Evolution, 8*(10), 1360-1373. doi:10.1111/2041-210x.12775
945 Patton, J. C., Gallaway, B. J., Fechhelm, R. G., & Cronin, M. A. (1997). Genetic variation of
946 microsatellite and mitochondrial DNA markers in broad whitefish (Coregonus nasus) in the
947 Colville and Sagavanirktok rivers in northern Alaska. *Canadian Journal of Fisheries and Aquatic*
948 *Sciences, 54*(7), 1548-1556.
949 Pearse, D. E., Barson, N. J., Nome, T., Gao, G., Campbell, M. A., Abadía-Cardoso, A., . . . Lien, S.
950 (2018). Sex-dependent dominance maintains migration supergene in rainbow trout. *bioRxiv*,
951 504621. doi:10.1101/504621
952 Perreault-Payette, A., Muir, A. M., Goetz, F., Perrier, C., Normandeau, E., Sirois, P., & Bernatchez, L.
953 (2017). Investigating the extent of parallelism in morphological and genomic divergence among
954 lake trout ecotypes in Lake Superior. *Molecular Ecology, 26*(6), 1477-1497.
955 doi:10.1111/mec.14018
956 Piette-Lauzière, G., Bell, A. H., Ridgway, M. S., & Turgeon, J. (2019). Evolution and diversity of two
957 cisco forms in an outlet of glacial Lake Algonquin. *Ecology and evolution, 0*(0).
958 doi:10.1002/ece3.5496
959 Præbel, K., Knudsen, R., Siwertsson, A., Karhunen, M., Kahilainen, K. K., Ovaskainen, O., . . .
960 Amundsen, P. A. (2013). Ecological speciation in postglacial European whitefish: rapid adaptive
961 radiations into the littoral, pelagic, and profundal lake habitats. *Ecology and evolution, 3*(15),
962 4970-4986.

963 Prince, D. J., O'Rourke, S. M., Thompson, T. Q., Ali, O. A., Lyman, H. S., Saglam, I. K., . . . Miller, M.
964 R. (2017). The evolutionary basis of premature migration in Pacific salmon highlights the utility
965 of genomics for informing conservation. *Science Advances, 3*(8), e1603198.
966 doi:10.1126/sciadv.1603198
967 Reed, K. M., Dorschner, M. O., Todd, T. N., & Phillips, R. B. (1998). Sequence analysis of the
968 mitochondrial DNA control region of ciscoes (genus Coregonus): taxonomic implications for the
969 Great Lakes species flock. *Molecular Ecology, 7*(9), 1091-1096.
970 Renaut, S., Nolte, A. W., & Bernatchez, L. (2009). Gene Expression Divergence and Hybrid
971 Misexpression between Lake Whitefish Species Pairs (Coregonus spp. Salmonidae). *Molecular*
972 *Biology and Evolution, 26*(4), 925-936. doi:10.1093/molbev/msp017
973 Rogers, S. M., & Bernatchez, L. (2006). The genetic basis of intrinsic and extrinsic post-zygotic
974 reproductive isolation jointly promoting speciation in the lake whitefish species complex
975 (Coregonus clupeaformis). *Journal of Evolutionary Biology, 19*(6), 1979-1994.
976 doi:10.1111/j.1420-9101.2006.01150.x
977 Rogers, S. M., & Bernatchez, L. (2007). The Genetic Architecture of Ecological Speciation and the
978 Association with Signatures of Selection in Natural Lake Whitefish (Coregonus sp. Salmonidae)
979 Species Pairs. *Molecular Biology and Evolution, 24*(6), 1423-1438. doi:10.1093/molbev/msm066
980 Rogers, S. M., Marchand, M. H., & Bernatchez, L. (2004). Isolation, characterization and cross-salmonid
981 amplification of 31 microsatellite loci in the lake whitefish (Coregonus clupeaformis, Mitchill).
982 *Molecular Ecology Notes, 4*(1), 89-92.
983 Rougeux, C., Gagnaire, P.-A., Praebel, K., Seehausen, O., & Bernatchez, L. (2019). Polygenic selection
984 drives the evolution of convergent transcriptomic landscapes across continents within a Nearctic
985 sister-species complex. *bioRxiv*, 311464. doi:10.1101/311464
986 Rousset, F. (2008). genepop'007: a complete re-implementation of the genepop software for Windows
987 and Linux. *Molecular ecology resources, 8*(1), 103-106. doi:10.1111/j.1471-8286.2007.01931.x
988 Ryder, O. A. (1986). Species conservation and systematics: the dilemma of subspecies. *Trends in Ecology*
989 *& Evolution, 1*(1), 9-10. doi:https://doi.org/10.1016/0169-5347(86)90059-5
990 Samuk, K., Owens, G. L., Delmore, K. E., Miller, S. E., Rennison, D. J., & Schluter, D. (2017). Gene
991 flow and selection interact to promote adaptive divergence in regions of low recombination.
992 *Molecular Ecology, 26*(17), 4378-4390. doi:10.1111/mec.14226
993 Schluter, D. (1996). Ecological speciation in postglacial fishes. *Philosophical Transactions of the Royal*
994 *Society of London. Series B: Biological Sciences, 351*(1341), 807-814.
995 Schmidt, S. N., Harvey, C. J., & Vander Zanden, M. J. (2011). Historical and contemporary trophic niche
996 partitioning among Laurentian Great Lakes coregonines. *Ecological Applications, 21*(3), 888-
997 896.
998 Schmidt, S. N., Vander Zanden, M. J., & Kitchell, J. F. (2009). Long-term food web change in Lake
999 Superior. *Canadian Journal of Fisheries and Aquatic Sciences, 66*(12), 2118-2129.
1000 Schwartz, M. K., Luikart, G., & Waples, R. S. (2006). Genetic monitoring as a promising tool for
1001 conservation and management. *Trends in Ecology & Evolution, 22*(1), 25-33.
1002 Scott, W., & Crossman, E. (1998). Freshwater fishes of Canada Oakville. *Ontario.: Galt House*
1003 *Publishing*.
1004 Seehausen, O. (2006). African Cichlid Fish: A Model System in Adaptive Radiation Research.
1005 *Proceedings: Biological Sciences, 273*(1597), 1987-1998.
1006 Shields, B. A., Guise, K. S., & Underhill, J. C. (1990). Chromosomal and mitochondrial DNA
1007 characterization of a population of dwarf cisco (Coregonus artedii) in Minnesota. *Canadian*
1008 *Journal of Fisheries and Aquatic Sciences, 47*(8), 1562-1569.
1009 Sierszen, M. E., Hrabik, T. R., Stockwell, J. D., Cotter, A. M., Hoffman, J. C., & Yule, D. L. (2014).
1010 Depth gradients in food-web processes linking habitats in large lakes: L ake S uperior as an
1011 exemplar ecosystem. *Freshwater Biology, 59*(10), 2122-2136.
1012 Smith, S. H. (1968). Species succession and fishery exploitation in the Great Lakes. *Journal of the*
1013 *Fisheries Board of Canada, 25*(4), 667-693.

Smith, S. H. (1970). Species interactions of the alewife in the Great Lakes. *Transactions of the American Fisheries Society, 99*(4), 754-765.

Soria-Carrasco, V., Gompert, Z., Comeault, A. A., Farkas, T. E., Parchman, T. L., Johnston, J. S., . . . Nosil, P. (2014). Stick Insect Genomes Reveal Natural Selection's Role in Parallel Speciation. *Science, 344*(6185), 738. doi:10.1126/science.1252136

Stockwell, J. D., Hrabik, T. R., Jensen, O. P., Yule, D. L., & Balge, M. (2010). Empirical evaluation of predator-driven diel vertical migration in Lake Superior. *Canadian Journal of Fisheries and Aquatic Sciences, 67*(3), 473-485.

Stott, W., MacDougall, T., Roseman, E. F., Lenart, S., Chiotti, J., Yule, D. L., & Boase, J. (in press). *Genetic Species Identification of Coregonines from Juvenile Assessment and Commercial Fisheries in Lake Erie, Detroit River and St. Clair River.* Paper presented at the 13th International Coregonid Symposium, Bayfield, Wiscosin, USA.

Svärdson, G. (1949). The coregonid problem. I. Some general aspects of the problem. *Rep. Inst. Freshw. Res. Drottningholm, 29*, 89-101.

Svärdson, G. (1979). Speciation of Scandinavian Coregonus [fresh water fish, Sweden]. *Report-Institute of Freshwater Research*.

Svärdson, G. (1998). Postglacial dispersal and reticulate evolution of Nordic coregonids. *Nordic Journal of Freshwater Research, 74*, 3-32.

Todd, T. N. (1998). Environmental modification of gillraker number in coregonine fishes. *Biology and management of Coregonid fishes-1996*, 305-315.

Todd, T. N., Smith, G. R., & Cable, L. E. (1981). Environmental and genetic contributions to morphological differentiation in ciscoes (Coregoninae) of the Great Lakes. *Canadian Journal of Fisheries and Aquatic Sciences, 38*(1), 59-67.

Turgeon, J., & Bernatchez, L. (2003). Reticulate evolution and phenotypic diversity in North American ciscoes, Coregonus ssp.(Teleostei: Salmonidae): implications for the conservation of an evolutionary legacy. *Conservation Genetics, 4*(1), 67-81.

Turgeon, J., Estoup, A., & Bernatchez, L. (1999). Species flock in the North American Great Lakes: molecular ecology of lake Nipigon ciscoes (Teleostei: coregonidae: coregonus). *Evolution, 53*(6), 1857-1871.

Turgeon, J., Reid, S. M., Bourret, A., Pratt, T. C., Reist, J. D., Muir, A. M., & Howland, K. L. (2016). Morphological and genetic variation in Cisco (Coregonus artedi) and Shortjaw Cisco (C. zenithicus): multiple origins of Shortjaw Cisco in inland lakes require a lake-specific conservation approach. *Conservation Genetics, 17*(1), 45-56.

Vendrami, D. L. J., Telesca, L., Weigand, H., Weiss, M., Fawcett, K., Lehman, K., . . . Hoffman, J. I. (2017). RAD sequencing resolves fine-scale population structure in a benthic invertebrate: implications for understanding phenotypic plasticity. *Royal Society Open Science, 4*(2), 160548. doi:10.1098/rsos.160548

Via, S. (2012). Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1587), 451-460. doi:10.1098/rstb.2011.0260

Via, S., & West, J. (2008). The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology, 17*(19), 4334-4345. doi:10.1111/j.1365-294X.2008.03921.x

Vuorinen, J. A., Bodaly, R. A., Reist, J. D., Bernatchez, L., & Dodson, J. J. (1993). Genetic and Morphological Differentiation between Dwarf and Normal Size Forms of Lake Whitefish (Coregonus clupeafotmis) in Como Lake, Ontario. *Canadian Journal of Fisheries and Aquatic Sciences, 50*(1), 210-216. doi:10.1139/f93-023

Wagner, C. E., Keller, I., Wittwer, S., Selz, O. M., Mwaiko, S., Greuter, L., . . . Seehausen, O. (2013). Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology, 22*(3), 787-798. doi:10.1111/mec.12023

Waples, R. K., Larson, W. A., & Waples, R. S. (2016). Estimating contemporary effective population size in non-model species using linkage disequilibrium across thousands of loci. *Heredity, 117*(4), 233.

Waples, R. S. (2006). A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci*. *Conservation Genetics, 7*(2), 167-184. doi:10.1007/s10592-005-9100-y

Waples, R. S., Antao, T., & Luikart, G. (2014). Effects of overlapping generations on linkage disequilibrium estimates of effective population size. *Genetics, 197*(2), 769-780.

Waples, R. S., & Do, C. (2010). Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. *Evolutionary Applications, 3*(3), 244-262.

Waples, R. S., & Lindley, S. T. (2018). Genomics and conservation units: The genetic basis of adult migration timing in Pacific salmonids. *Evolutionary Applications, 11*(9), 1518-1526.

Weigel, D., Adams, J., A Jepson, M., Waits, L., Caudill, C., & Weigel, D. (2018). Introgressive Hybridization Between Native and Hatchery-origin Non-native Steelhead (Oncorhynchus mykiss). *Aquatic Conservation Marine and Freshwater Ecosystems, 29*. doi:10.1002/aqc.3028

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution, 38*(6), 1358-1370. doi:10.2307/2408641

Wells, L., & McClain, A. (1973). Lake Michigan: man's effects on native fish stocks and other biota Great Lakes Fishery Commission Technical Report No. 20.

Whiteley, A. R., Persaud, K. N., Derome, N., Montgomerie, R., & Bernatchez, L. (2009). Reduced sperm performance in backcross hybrids between species pairs of whitefish (Coregonus clupeaformis). *Canadian Journal of Zoology, 87*(7), 566-572. doi:10.1139/Z09-042

Wilson, C. C., Stott, W., Miller, L., D'Amelio, S., Jennings, M. J., & Cooper, A. M. (2008). Conservation Genetics of Lake Superior Brook Trout: Issues, Questions, and Directions. *North American Journal of Fisheries Management, 28*(4), 1307-1320. doi:10.1577/M05-190.1

Yule, D., Stockwell, J., Evrard, L., Cholwek, G., & Cullis, K. (2006). Comparison of commercial landings of cisco to acoustic estimates of abundance in Thunder Bay and Black Bay, Ontario. *US Geological Survey, Great Lakes Science Center, Ann Arbor, Michigan*.

Yule, D. L., Addison, P. A., Evrard, L. M., Cullis, K. I., & Cholwek, G. A. (2009). 2008 spawning cisco investigations in the Canadian waters of Lake Superior. In *US Geological Survey Technical Report 4-08*